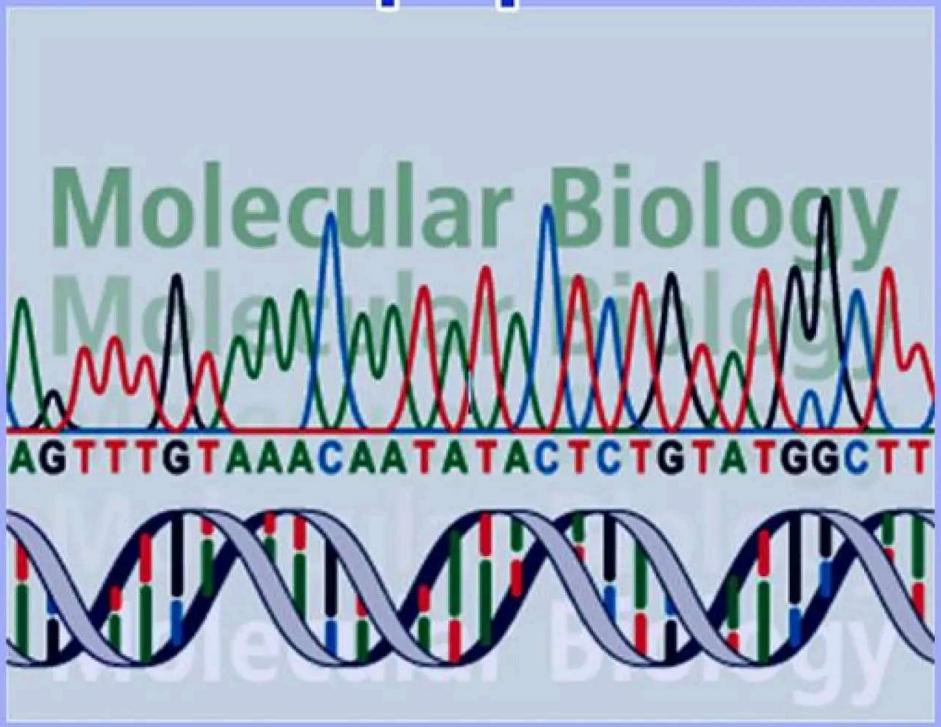
Nguyễn Hoàng Lộc (Chủ biên)
Trần Thị Lệ - Hà Thị Minh Thi

Giáo trình

Sinh học phân tử



NHÀ XUẤT BẢN ĐẠI HỌC HUẾ

Lời nói đầu

Sinh học phân tử là khoa học nghiên cứu các hiện tượng sống ở mức độ phân tử. Phạm vi nghiên cứu của môn học này có phần trùng lặp với một số môn học khác trong sinh học đặc biệt là di truyền học và hóa sinh học. Sinh học phân tử chủ yếu tập trung nghiên cứu mối tương tác giữa các hệ thống cấu trúc khác nhau trong tế bào, bao gồm mối quan hệ qua lại giữa quá trình tổng hợp của DNA, RNA và protein và tìm hiểu cách thức điều hòa các mối tương tác này.

Hiện nay, sinh học phân tử và sinh học tế bào được xem là nền tảng quan trọng của công nghệ sinh học. Nhờ phát triển các công cụ cơ bản của sinh học phân tử như các enzyme cắt hạn chế, DNA ligase, các vector tạo dòng, lai phân tử, kỹ thuật PCR... sinh học phân tử ngày càng đạt nhiều thành tựu ứng dụng quan trọng.

Giáo trình sinh học phân tử này cung cấp những kiến thức cơ bản cho sinh viên với các nội dung chính sau:

- Cấu trúc và chức năng của gen
- Cấu trúc genome
- Các quá trình tái bản, phiên mã và dịch mã của nguyên liệu di truyền
- Điều hòa biểu hiện gen
- Sửa chữa và bảo vệ gen
- Tái tổ hợp và chuyển gen

Do mới được xuất bản lần đầu nên giáo trình này khó tránh khỏi thiếu sót hoặc chưa đáp ứng được yêu cầu bạn đọc. Vì thế, chúng tôi mong nhận được nhiều ý kiến đóng góp để lần xuất bản sau được hoàn thiện hơn.

Chúng tôi chân thành cảm ơn Quỹ Nâng cao chất lượng-Dự án Giáo dục đại học đã hỗ trợ chúng tôi biên soạn giáo trình này, PGS. TS. Nông Văn Hải đã đọc bản thảo và góp nhiều ý kiến quý báu.

Các tác giả

Chương 1

Các đại phân tử sinh học

I. Nucleic acid

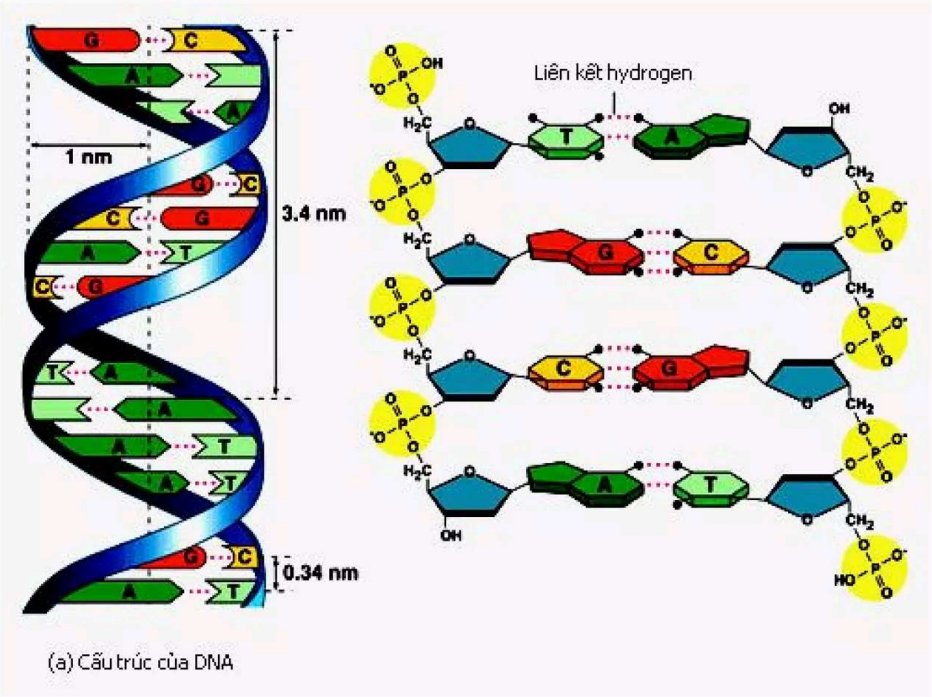
Nucleic acid, vật chất mang thông tin di truyền của các hệ thống sống, là một polymer hình thành từ các monomer là nucleotide. Mỗi nucleotide gồm ba thành phần: nhóm phosphate, đường pentose (đường 5 carbon) và một nitrogen base. Các nitrogen base thuộc hai nhóm: các purine gồm adenine (A) và guanine (G), các pyrimidine gồm thymine (T), cytosine (C) và uracil (U). Các nucleotide được nối với nhau bằng liên kết phosphodiester tạo thành chuỗi dài.

Nucleic acid gồm hai loại phân tử có cấu tạo rất giống nhau là deoxyribonucleic acid (DNA) và ribonucleic acid (RNA).

1. Deoxyribonucleic acid

Phân tử DNA là một chuỗi xoắn kép gồm hai sợi đơn. Mỗi sợi đơn là một chuỗi nucleotide (Hình 1.1). Mỗi nucleotide gồm ba thành phần: nhóm phosphate, đường deoxyribose và một trong bốn base (A, C, G và T) (Hình 1.2). Hai sợi đơn kết hợp với nhau nhờ các liên kết hydrogen hình thành giữa các base bổ sung nằm trên hai sợi: A bổ sung cho T và C bổ sung cho G. Mỗi sợi đơn có một trình tự định hướng với một đầu 5'phosphate tự do, đầu kia là 3' hydroxyl tự do (quy ước là 5'→3'). Hướng của hai sợi đơn trong chuỗi xoắn kép ngược nhau, nên được gọi là hai sợi đối song.

Những phân tích cấu trúc hiện đại đã cho thấy cấu trúc của DNA không phải luôn luôn tương ứng với dạng được gọi là B mà Watson và Crick đã đưa ra. Do sự tác động của các hợp chất có khối lượng nhỏ hoặc protein, dạng B có thể chuyển sang dạng A (nén nhiều hơn) hoặc là dạng Z (xoắn trái). Chúng có thể tự gấp lại hoặc xoắn mạnh, ví dụ một sợi đôi DNA có độ dài là 20 cm được nén trong một chromosome có kích thước là 5 μm.

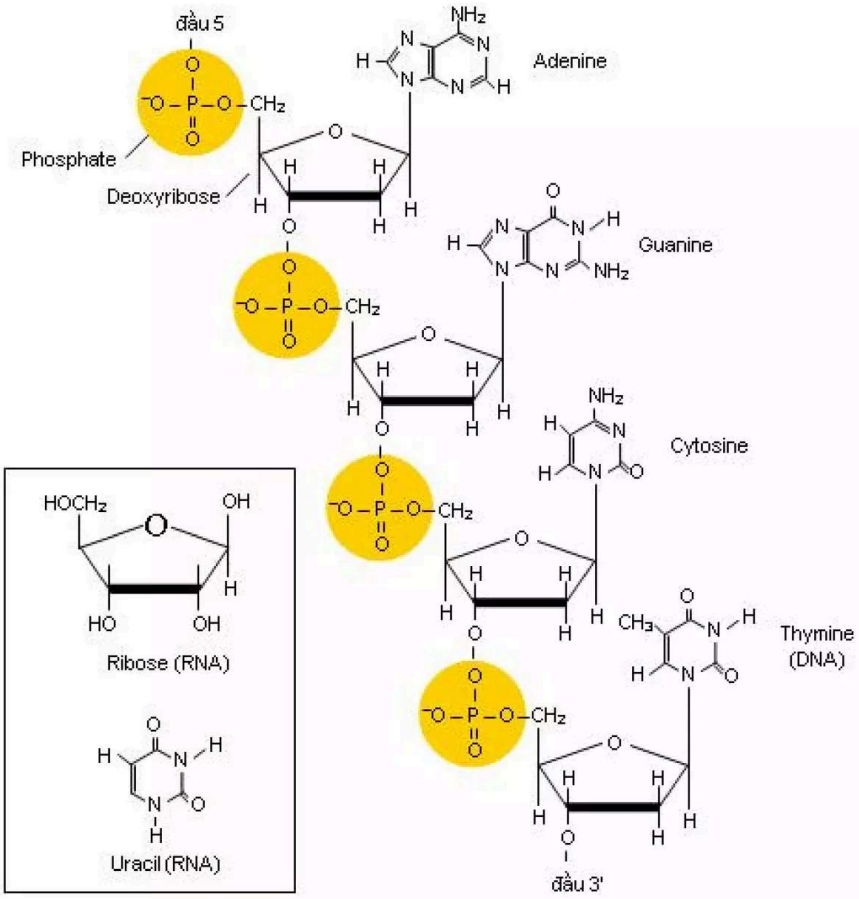


Hình 1.1. Chuỗi xoắn kép của DNA

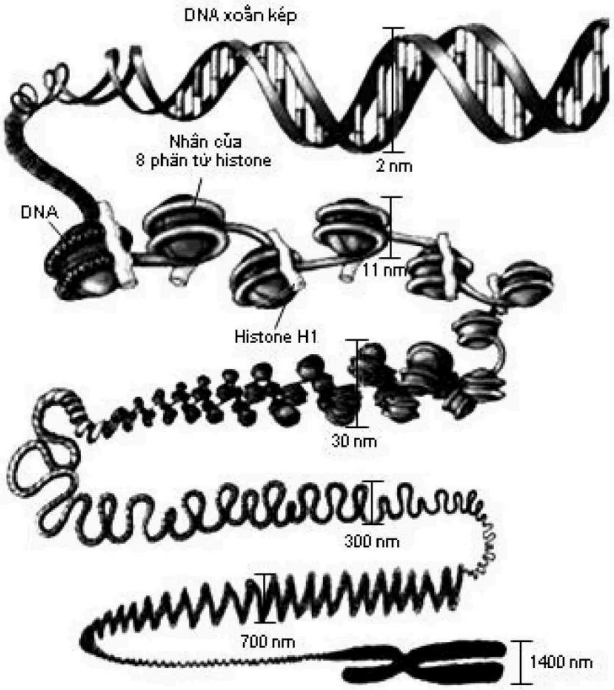
Phân tử DNA trong nhiễm sắc thể của sinh vật eukaryote ở dạng mạch thẳng, còn ở phần lớn tế bào prokaryote (vi khuẩn) phân tử DNA lại có dạng mạch vòng. Tuy nhiên, dù ở dạng nào thì các phân tử DNA này đều tồn tại theo kiểu cuộn chặt. Trong tế bào eukaryote, DNA kết hợp chặt chẽ với các protein là histone.

DNA eukaryote có kích thước rất lớn (Ví dụ: DNA ở người có thể dài đến 1 m) nên vấn đề đặt ra là phân tử này phải được nén như thế nào trong một thể tích rất hạn chế của nhân. Việc nén được thực hiện ở nhiều mức độ, mức độ thấp nhất là nucleosome và mức độ cao nhất là cấu trúc nhiễm sắc chất. Thật vậy, đường kính của chuỗi xoắn DNA chỉ là 20 Å, trong khi sợi nhiễm sắc chất quan sát dưới kính hiển vi điện tử có đường kính 100 Å, đôi khi đạt 300 Å. Điều này chứng tỏ phân tử DNA tham gia hình thành những cấu trúc phức tạp hơn (Hình 1.3).

Sợi nhiễm sắc chất có đường kính 100 Å là một chuỗi chứa nhiều nucleosome. Đó là những cấu trúc hình thành từ một sợi DNA quấn quanh một lõi gồm 8 phân tử histone (mức độ tổ chức cao nhất của DNA). Sợi có đường kính 100 Å này có cấu trúc phức tạp hơn sợi có đường kính 300 Å. Trong nhân tế bào, các sợi vừa kể trên kết hợp chặt chẽ với nhiều protein khác nhau và cả với các RNA tạo thành nhiễm sắc chất, mức độ tổ chức cao nhất của DNA.



Hình 1.2. Cấu trúc các nucleotide điển hình



Hình 1.3. Cấu trúc nucleosome và nhiễm sắc thể. Phân tử DNA được cuộn lại trên nhiễm sắc thể làm cho chiều dài ngắn lại hơn 50.000 lần.

Các DNA ở eukaryote có đặc điểm khác với DNA prokaryote. Toàn bộ phân tử DNA prokaryote đều mang thông tin mã hóa cho các protein trong khi đó DNA của eukaryote bao gồm những trình tự mã hóa (các exon) xen kẽ với những trình tự không mã hóa (intron). Các trình tự mã hóa ở eukaryote chìm ngập trong một khối lớn DNA mà cho đến nay vẫn chưa rõ tác dụng được gọi là "DNA rác" (junk DNA). Tùy theo mức độ hiện diện của chúng trong nhân, các trình tự DNA được chia làm ba loại:

- **Các trình tự lặp lại nhiều lần.** Ví dụ: ở động vật có vú các trình tự này chiếm 10-15% genome (hệ gen). Đó là những trình tự DNA ngắn (10-200 kb), không mã hóa, thường tập trung ở những vùng chuyên biệt trên nhiễm sắc thể như ở vùng tâm động (trình tự CEN) hay ở đầu các nhiễm sắc thể (trình tự TEL). Chức năng của các trình tự này chưa rõ, có thể chúng tham gia vào quá trình di chuyển DNA trên thoi vô sắc (trình tự CEN) hoặc vào quá trình sao chép toàn bộ phần DNA nằm ở đầu mút nhiễm sắc thể (trình tự TEL).

- **Các trình tự có số lần lặp lại trung bình.** Ví dụ: ở genome người các trình tự này chiếm 25-40 %. Chúng đa dạng hơn và có kích thước lớn hơn (100-1.000 kb) các trình tự lặp lại nhiều lần. Các trình tự này phân bố trên toàn bộ genome. Chúng có thể là

những trình tự không mã hóa mà cũng có thể là những trình tự mã hóa cho rRNA, tRNA và 5S RNA.

- **Các trình tự duy nhất.** Là các gen mã hóa cho các protein, có trình tự đặc trưng cho từng gen.

Một đặc điểm của phân tử DNA có ý nghĩa rất quan trọng và được sử dụng vào phương pháp lai phân tử, đó là khả năng biến tính và hồi tính. Biến tính là hiện tượng hai sợi đơn của phân tử DNA tách rời nhau khi các liên kết hydrogen giữa các base bổ sung nằm trên hai sợi bị đứt do các tác nhân hóa học (dung dịch kiềm, formamide, urea) hay do tác nhân vật lý (nhiệt). Sau đó, nếu điều chỉnh nhiệt độ và nồng độ muối thích hợp, các sợi đơn có thể bắt cặp trở lại theo nguyên tắc bổ sung, để hình thành phân tử DNA ban đầu, đó là sự hồi tính.

2. Ribonucleic acid

Phân tử RNA có cấu tạo tương tự DNA ngoại trừ ba điểm khác biệt sau:

- Phân tử RNA là chuỗi đơn.
- Đường pentose của phân tử RNA là ribose thay vì deoxyribose.
- Thymine (T), một trong bốn loại base hình thành nên phân tử DNA, được thay thế bằng uracil (U) trong phân tử RNA.

Cấu trúc và chức năng của RNA có sự biến đổi rõ rệt. Về cơ bản RNA chỉ là chất mang thông tin di truyền ở virus, sau đó người ta chứng minh rằng nó không những đóng vai trò cơ bản ở việc chuyển thông tin di truyền mà còn có vai trò cấu trúc khi tạo nên phức hệ RNA-protein.

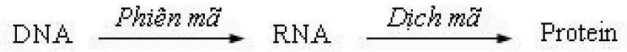
Theo một lý thuyết tiến hóa mà đại diện là Eigen, RNA là chất mang thông tin di truyền, thành viên trung gian của sự biểu hiện gen, thành phần cấu tạo và là chất xúc tác. Nhóm OH ở vị trí thứ hai của ribose cần thiết cho đa chức năng làm nhiễu loạn sự tạo thành sợi đôi, qua đó làm tăng độ không bền vững của liên kết phosphodieste.

Trong tế bào có ba loại RNA chính, có các chức năng khác nhau:

2.1. Các RNA thông tin (mRNA)

mRNA là bản sao của những trình tự nhất định trên phân tử DNA, có vai trò trung tâm là chuyển thông tin mã hóa trên phân tử DNA đến bộ máy giải mã thành phân tử protein tương ứng. Các RNA có cấu trúc đa dạng, kích thước nhỏ hơn so với DNA vì chỉ chứa thông tin mã hóa cho một hoặc vài protein và chỉ chiếm khoảng 2-5% tổng số RNA trong tế bào.

Quá trình chuyển thông tin được thể hiện như sau:



Ở *E. coli*, kích thước trung bình của một phân tử mRNA khoảng 1,2 kb.

2.2. RNA vận chuyển (tRNA)

tRNA làm nhiệm vụ vận chuyển các amino acid hoạt hóa đến ribosome để tổng hợp protein từ các mRNA tương ứng. Có ít nhất một loại tRNA cho một loại amino acid. tRNA vận chuyển chứa khoảng 75 nucleotide (có khối lượng khoảng 25 kDa), là phân tử RNA nhỏ nhất. Các tRNA có cấu trúc dạng cỏ ba lá. Cấu trúc này được ổn định nhờ các liên kết bổ sung hiện diện ở nhiều vùng của phân tử tRNA. Hai vị trí không có liên kết bổ sung đóng vai trò đặc biệt quan trọng đối với chức năng của tRNA:

- Trình tự anticodon gồm ba nucleotide.
- Trình tự CCA, có khả năng liên kết cộng hóa trị với một amino acid đặc trưng.

2.3. RNA ribosome (rRNA)

rRNA là thành phần cơ bản của ribosome, đóng vai trò xúc tác và cấu trúc trong tổng hợp protein.

Tùy theo hệ số lắng rRNA được chia thành nhiều loại: ở eukaryote có 28S; 18S; 5,8S và 5S rRNA; còn các rRNA ở *E. coli* có ba loại: 23S, 16S và 5S.

rRNA chiếm nhiều nhất trong ba loại RNA (80% tổng số RNA tế bào), tiếp đến là tRNA khoảng 16% và mRNA chỉ khoảng 2%. Ngoài ra, tế bào sinh vật eukaryote còn chứa những phân tử RNA kích thước nhỏ của nhân (small nuclear, snRNA) chiếm khoảng <1% tham gia vào ghép nối các exon. Ribosome là những phân tử cần thiết cho sự tổng hợp protein, ribosome của mọi tế bào đều gồm một tiểu đơn vị nhỏ và một tiểu đơn vị lớn. Mỗi tiểu đơn vị có mang nhiều protein và rRNA (trong đó rRNA là thành phần chủ yếu chiếm khoảng 65%) có kích thước khác nhau. Người ta cũng thấy ribosome trong ty thể, ở đó có sự tổng hợp một số protein ty thể.

Bảng 1.1. Các phân tử RNA trong *E. coli*

Loại	Tổng số tương đối (%)	Hệ số lắng (S)[1]	Khối lượng phân tử (kDa)	Số lượng nucleotide
rRNA	80	23 16 5	$1,2 \times 10^3$ $0,55 \times 10^3$ $3,6 \times 10^1$	3700 1700 120
tRNA	15	4	$2,5 \times 10^1$	75
mRNA	5		Không đồng nhất	

2.3.1. Ribosome của prokaryote

Tế bào được nghiên cứu về ribosome nhiều nhất là *E. coli.* Ribosome (70S) của *E. coli* gồm hai tiểu đơn vị: tiểu đơn vị nhỏ (30S) và tiểu đơn vị lớn (50S). Căn cứ vào hệ số lắng, người ta phân biệt ba loại rRNA: 23S rRNA, 16S rRNA và 5S rRNA.

- Tiểu đơn vị 30S chứa: 1 phân tử 16S rRNA (có 1540 nu) và 21 ribosomal protein khác nhau.

- Tiểu đơn vị 50S chứa: 1 phân tử 5S rRNA (có 120 nu), 1 phân tử 23S rRNA (có 2900 nu) và 34 ribosomal protein.

Hai tiểu đơn vị nhỏ và lớn khi kết hợp với nhau sẽ tạo ra một rãnh ở chỗ tiếp giáp của chúng để cho mRNA đi qua.

2.3.2. Ribosome của eukaryote

Ribosome của eukaryote (80S) lớn hơn ribosome của prokaryote cũng bao gồm hai tiểu đơn vị: tiểu đơn vị nhỏ (40S) và tiểu đơn vị lớn (60S).

- Tiểu đơn vị 40S chứa: 1 phân tử 18S rRNA (có 1900 nu) và 33 ribosomal protein.

- Tiểu đơn vị 60S chứa: 3 phân tử rRNA (5S; 5,8S và 28S) và 49 ribosomal protein.

Tóm lại, tất cả RNA trong tế bào đều được tổng hợp nhờ enzyme RNA polymerase. Enzyme này đòi hỏi những thành phần sau đây:

- Một khuôn mẫu, thường là DNA sợi đôi.

- Tiền chất hoạt hóa: Bốn loại ribonucleoside triphosphate: ATP, GTP, UTP và CTP.

Sinh tổng hợp RNA giống DNA ở một số điểm, thứ nhất hướng tổng hợp là 5'→3', thứ hai là cơ chế kéo dài giống nhau: nhóm 3'-OH ở đầu cuối của chuỗi tổng hợp là vị trí

gắn kết của nucleoside triphosphate tiếp theo. Thứ ba, sự tổng hợp xảy ra do thủy phân pyrophosphate.

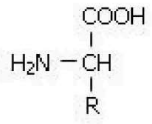
Tuy nhiên, khác với DNA là RNA không đòi hỏi mồi (primer). Ngoài ra, RNA polymerase không có hoạt tính nuclease để sửa chữa khi các nucleotide bị gắn nhầm.

Cả ba loại RNA trong tế bào được tổng hợp trong *E. coli* nhờ một loại RNA polymerase. Ở động vật có vú, các RNA khác nhau được tổng hợp bằng các loại RNA polymerase khác nhau.

II. Protein

1. Cấu trúc của protein

Amino acid là đơn vị cơ sở (monomer) cấu thành protein. Tất cả 20 amino acid có mặt trong protein đều được xây dựng theo một kiểu mẫu chung:



Công thức tổng quát của L-α-amino acid

Trong đó, gốc R (mạch bên) cũng là phần khác duy nhất giữa 20 loại amino acid, quy định tính chất của từng loại.

Nhóm amine (NH_2) đính ở nguyên tử C_2, theo tên cũ là nguyên tử C_α. Vì vậy, người ta gọi là nhóm α-amine. Các amino acid tồn tại chủ yếu trong tự nhiên có nhóm amine đứng ở bên trái trục, được gọi là amino acid dạng L. Dạng D-amino acid chỉ tồn tại riêng biệt, ví dụ trong thành tế bào vi khuẩn.

Các amino acid riêng biệt có những đặc tính khác nhau là do gốc R của chúng. Những amino acid trung tính có một nhóm amine và một nhóm carboxyl. Những protein chứa nhiều amino acid trung tính là những protein trung tính. Khi chiều dài gốc R tăng sẽ hình thành đặc tính kỵ nước. Những protein có chứa nhiều amino acid như valine, leucine, isoleucine có tính chất đặc trưng là kỵ nước. Những amino acid có tính acid trong phần gốc có một nhóm carboxyl. Protein chứa nhiều amino acid có tính acid là những protein acid. Tương tự như vậy đối với protein chủ yếu được hình thành bởi những amino acid có tính kiềm là những protein kiềm. Phần gốc R của amino acid có ý nghĩa quyết định đối với đặc tính của protein mà chúng tạo nên. Điều này không những có ý nghĩa đối với tính chất hóa học mà cả cấu trúc của protein.

Thủy phân hoàn toàn protein, thu được chủ yếu các L-α-amino acid. Mặc dù protein rất đa dạng nhưng hầu hết chúng đều được cấu tạo từ 20 L-α-amino acid. Dựa vào đặc tính của gốc R, amino acid được chia làm bảy nhóm chính sau đây:

- **Amino acid trung tính mạch thẳng.** Bao gồm glycine, alanine, valine, leucine và isoleucine.

- **Các hydroxyl amino acid mạch thẳng.** Bao gồm serine và threonine.

- **Amino acid chứa lưu huỳnh mạch thẳng.** Bao gồm cysteine và methionine. Khi oxy hóa hai nhóm -SH của hai phân tử cysteine tạo thành cystine có chứa cầu (-S-S-).

- **Các amino acid acid và các amide.** Bao gồm aspartic acid và glutamic acid. Trong phân tử của chúng có một nhóm amine và hai nhóm carboxyl. Ở độ pH sinh lý (6-7), các amino acid này tích điện âm. Amine hóa nhóm carboxyl ở mạch bên của aspartate và glutamate tạo thành các amide tương ứng là asparagine và glutamine.

- **Các amino acid kiềm**. Bao gồm lysine và arginine.

- **Iminoacid.** Proline.

- **Các amino acid thơm và dị vòng.** Bao gồm phenylalanine, tyrosine và tryptophan. Do có chứa vòng thơm nên các amino acid này có một số phản ứng đặc trưng.

Các amino acid được nối với nhau bởi các liên kết peptide, liên kết này được hình thành do sự kết hợp nhóm amine của một amino acid với nhóm carboxyl của amino acid kế tiếp. Phản ứng kết hợp giải phóng một phân tử H_2O.

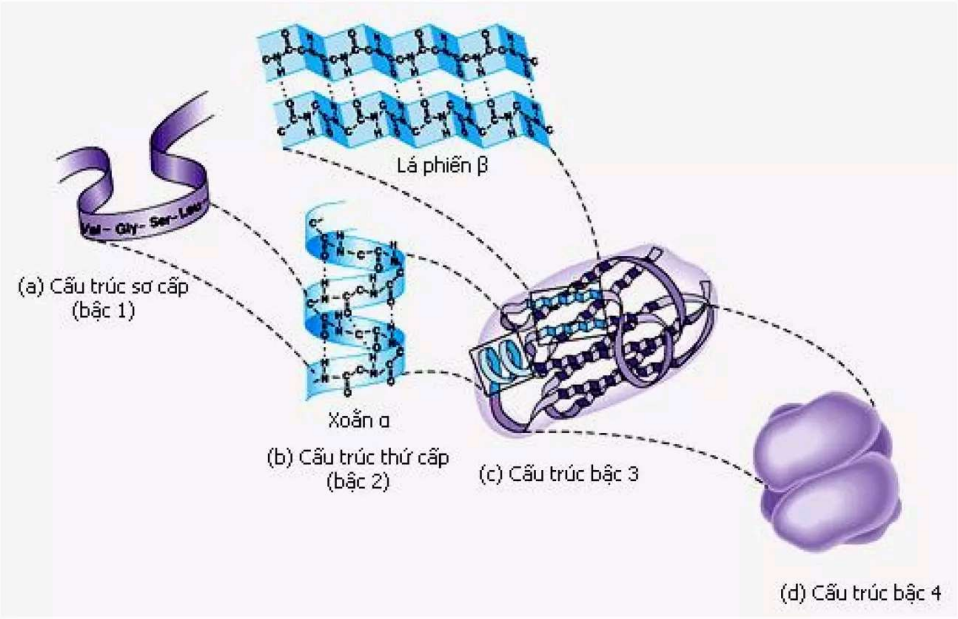
Peptide là một chuỗi nối tiếp nhiều amino acid (số lượng ít hơn 30). Với số lượng amino acid lớn hơn chuỗi được gọi là polypeptide. Mỗi polypeptide có hai đầu tận cùng, một đầu mang nhóm amine tự do, đầu kia mang nhóm carboxyl tự do. Protein được dùng để chỉ đơn vị chức năng, nghĩa là một cấu trúc phức tạp trong không gian chứ không phải đơn thuần là một trình tự amino acid.

Chuỗi polypeptide có thể uốn thành cấu trúc hình gậy như trong các protein hình sợi hay cấu trúc khối cầu như trong các protein dạng cầu hay một cấu trúc gồm cả hai dạng trên. Một protein có thể được hình thành từ nhiều chuỗi polypeptide.

Người ta thường phân biệt cấu trúc của phân tử protein thành bốn bậc như sau (Hình 1.4):

- **Cấu trúc bậc 1**. Là trình tự sắp xếp các gốc amino acid trong chuỗi polypeptide. Cấu trúc này được giữ vững nhờ liên kết peptide (liên kết cộng hóa trị).

Vì mỗi một amino acid có gốc khác nhau, các gốc này có những đặc tính hóa học khác nhau, nên một chuỗi polypeptide ở các thời điểm khác nhau có những đặc tính hóa học rất khác nhau. Tuy nhiên, về tổng quát thì tất cả các chuỗi polypeptide được xây dựng một cách có hệ thống từ các nhóm nguyên tử CO, CH và NH. Việc xây dựng có hệ thống này là cơ sở để tạo nên cấu trúc bậc hai.



Hình 1.4. Các mức độ tổ chức của phân tử protein

- **Cấu trúc bậc 2.** Là tương tác không gian giữa các gốc amino acid ở gần nhau trong chuỗi polypeptide. Cấu trúc được bền vững chủ yếu nhờ liên kết hydrogen hình thành giữa các liên kết peptide ở kề gần nhau, cách nhau những khoảng xác định.

Cấu trúc bậc 2 của phân tử protein: xoắn α (α-helix), lá phiến β và xoắn collagen. Loại α-helix là sợi ở dạng xoắn ốc, cuộn xung quanh một trục, mỗi vòng xoắn có 3,6 gốc amino acid.

Những sợi collagen chạy song song tạo nên những bó sợi dai của gân. Collagen cũng có trong xương và trong các mô nối. Elastin là một protein, gồm những sợi protein tương đối ngắn, gắn kết với nhau nhờ liên kết cộng hóa trị. Những chuỗi polypeptide quay theo dạng xoắn ốc, tự duỗi xoắn khi có áp lực.

- Cấu trúc bậc 3. Là tương tác không gian giữa các gốc amino acid ở xa nhau trong chuỗi polypeptide, là dạng cuộn lại trong không gian của toàn chuỗi polypeptide.

Nhiều chuỗi polypeptide trong cơ thể sống tồn tại không phải ở dạng thẳng mà gấp khúc và qua đó tạo nên cấu trúc không gian ba chiều. Tuy nhiên, cấu trúc này hoàn toàn xác định, chủ yếu là do trình tự các amino acid và môi trường. Khi một chuỗi polypeptide tách ra khỏi ribosome sau khi tổng hợp và được đưa vào trong tế bào chất như là môi trường tạo hình thì nó sẽ hình thành nên cấu trúc tự nhiên rất nhanh, đặc biệt đối với cấu trúc hình cầu, đem lại cho protein những đặc tính sinh lý quan trọng. Có thể do chuyển động nhiệt của các chuỗi polypeptide mà các nhóm của các gốc amino acid tiếp xúc với

nhau, dẫn đến có thể kết hợp với nhau. Trong nhiều protein hình cầu có chứa các gốc cysteine, sự tạo thành các liên kết disulfite giữa các gốc cysteine ở xa nhau trong chuỗi polypeptide sẽ làm cho chuỗi bị cuộn lại đáng kể. Các liên kết khác, như liên kết Van der Waals, liên kết tĩnh điện, phân cực, kỵ nước và hydrogen giữa các mạch bên của các gốc amino acid đều tham gia làm bền vững cấu trúc bậc 3. Cấu trúc hình cầu của protein được gọi là cấu trúc bậc ba, đó chính là cấu trúc của enzyme.

- Cấu trúc bậc 4. Là tương tác không gian giữa các chuỗi của các phân tử protein gồm hai hay nhiều chuỗi polypeptide hình cầu. Mỗi chuỗi polypeptide này được gọi là một tiểu đơn vị (subunit). Sự kết hợp giữa các phân tử này lỏng lẻo và chủ yếu là do liên kết hydrogen và kỵ nước. Bằng cách này hai phân tử xác định có thể kết hợp với nhau tạo thành một dimer. Chẳng hạn: hemoglobin được tạo nên từ hai chuỗi α với mỗi chuỗi có 141 gốc amino acid và hai chuỗi β với mỗi chuỗi là 146 gốc amino acid.

Cấu trúc của một hoặc nhiều chuỗi polypeptide có ý nghĩa quan trọng đối với độ hòa tan và chức năng của chúng. Cấu trúc protein được hiểu là sự sắp xếp của những chuỗi riêng lẻ hoặc nhiều chuỗi. Chúng phụ thuộc nhiều vào độ pH của môi trường. Protein và chuỗi polypeptide hòa tan tốt khi những nhóm ưa nước hướng ra phía ngoài, nhóm kỵ nước hướng vào bên trong. Khi một protein thay đổi cấu trúc thì những nhóm kỵ nước quay ra ngoài, protein mất khả năng hòa tan trong nước, ví dụ trường hợp kết tủa không ở dạng tinh thể của protein sữa trong môi trường chua. Lactic acid được sản sinh do vi khuẩn làm giảm pH sữa, làm thay đổi protein sữa. Nhiều nhóm kỵ nước được hướng ra bên ngoài, protein mất khả năng tan trong nước. Vì vậy, việc thường xuyên duy trì giá trị pH trong tế bào chất rất quan trọng, vì chỉ có như vậy chức năng hoạt động của các enzyme trong tế bào chất mới được đảm bảo.

2. Chức năng của protein

Mỗi một hoạt động trong tế bào phụ thuộc vào một hoặc nhiều phân tử protein đặc hiệu. Một trong các cách phân loại protein là dựa vào chức năng sinh học của chúng. Bảng 1.2 tóm tắt sự phân loại protein theo chức năng và đưa ra một số ví dụ đại diện cho mỗi loại.

Bảng 1.2. Các chức năng sinh học của protein và một số ví dụ

Các nhóm chức năng	Ví dụ
Enzyme	Ribonuclease Trypsin Phosphofructokinase Alcohol dehydrogenase Catalase Malic enzyme
Protein điều khiển	Insulin Somatotropin Thyrotropin *lac* repressor NF1 (nuclear factor 1) Catabolite activator protein (CAP) AP1
Protein vận chuyển	Hemoglobin Serum albumin Glucose transporter
Protein dự trữ	Ovalbumin Casein Zein Phaseolin Ferritin
Protein vận động và co rút	Actin Myosin Tubulin Dynelin Kinesin
Protein cấu trúc	α-Keratin Collagen Elastin Fibroin Proteoglycans
Protein cấu trúc tạm thời (scaffold protein)	Grb 2 crk shc stat IRS-1
Protein bảo vệ	Immunoglobulins Thrombin Fibrinogen Antifreeze proteins Snake and bee venom proteins Diphtheria toxin Ricin

2.1. Chức năng enzyme

Phần lớn protein là enzyme. Hiện nay, có hơn 3.000 loại enzyme đã được biết. Enzyme là chất xúc tác sinh học có vai trò làm tăng tốc độ phản ứng. Mỗi một bước trong trao đổi chất đều được xúc tác bởi enzyme. Enzyme có thể làm tăng tốc độ phản ứng lên 10^{16} lần so với tốc độ phản ứng không xúc tác. Sự kết hợp giữa enzyme và cơ chất xảy ra ở vị trí hoạt động của enzyme.

2.2. Protein điều khiển

Một số protein không thực hiện bất kỳ sự biến đổi hóa học nào, tuy nhiên nó điều khiển các protein khác thực hiện chức năng sinh học, chẳng hạn insulin điều khiển nồng độ đường glucose trong máu. Đó là một protein nhỏ (5,7 kDa), gồm hai chuỗi polypeptide nối với nhau bằng các liên kết disulfite. Khi không đủ insulin thì sự tiếp nhận đường trong tế bào bị hạn chế. Vì vậy, mức đường trong máu tăng và dẫn đến sự thải đường mạnh mẽ qua nước tiểu (bệnh tiểu đường).

Một nhóm protein khác tham gia vào sự điều khiển biểu hiện gen. Những protein này có đặc tính là gắn vào những trình tự DNA hoặc để hoạt hóa hoặc ức chế sự phiên mã thông tin di truyền sang mRNA, ví dụ chất ức chế (repressor) đình chỉ sự phiên mã.

2.3. Protein vận chuyển

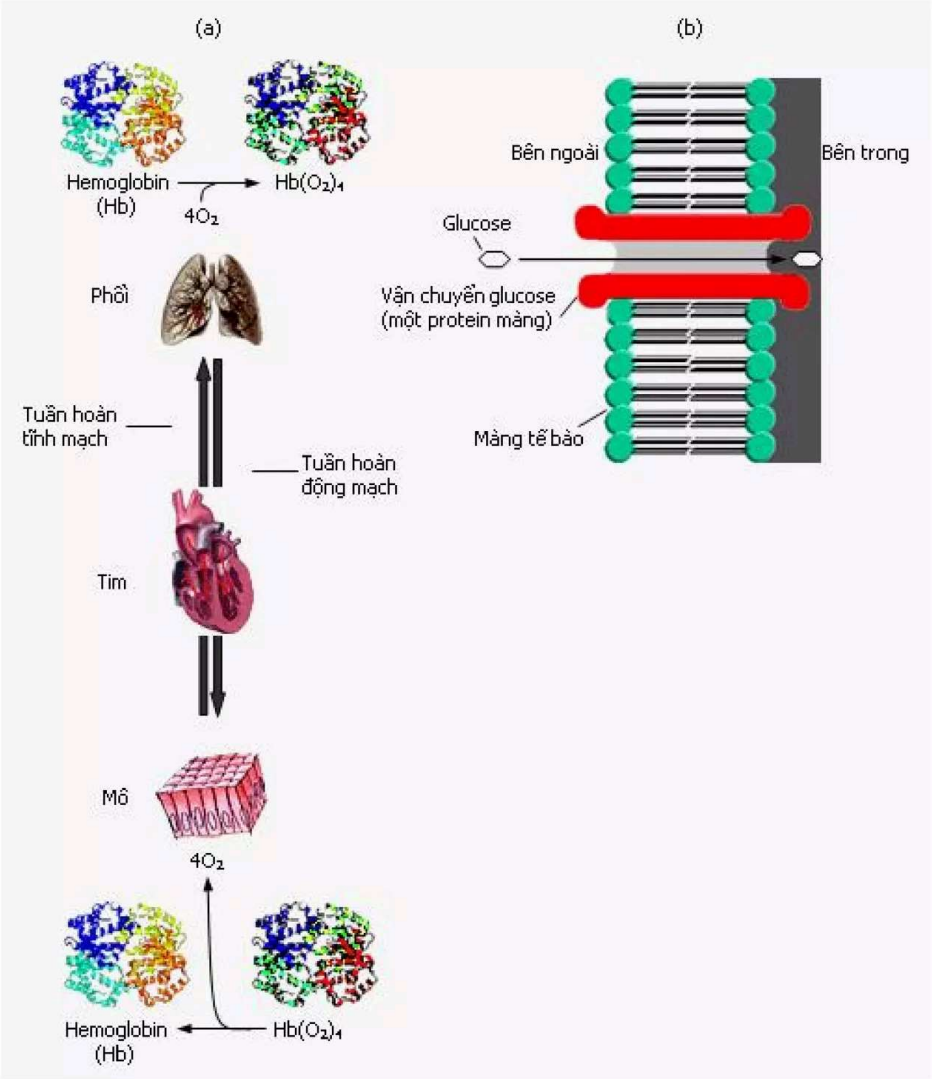
Làm nhiệm vụ vận chuyển chất đặc hiệu từ vị trí này sang vị trí khác, ví dụ vận chuyển O_2 từ phổi đến các mô do hemoglobin hoặc vận chuyển acid béo từ mô dự trữ đến các cơ quan khác nhờ protein trong máu là serum albumin.

Các chất được vận chuyển qua màng được thực hiện bằng các protein đặc hiệu, chẳng hạn vận chuyển glucose hoặc các amino acid qua màng (Hình 1.5).

2.4. Protein dự trữ

Các protein là nguồn cung cấp các chất cần thiết được gọi là protein dự trữ. Protein là polymer của các amino acid và nitrogen thường là yếu tố hạn chế cho sinh trưởng, nên cơ thể phải có protein dự trữ để cung cấp đầy đủ nitrogen khi cần. Chẳng hạn, ovalbumin là protein dự trữ trong lòng trắng trứng cung cấp đủ nitrogen cho phôi phát triển. Casein là protein sữa cung cấp nitrogen cho động vật có vú còn non. Hạt ở thực vật bậc cao cũng

chứa một lượng protein dự trữ lớn (khoảng 60%), cung cấp đủ nitrogen cho quá trình nảy mầm của hạt.



Hình 1.5. Hai kiểu vận chuyển cơ bản. (a): vận chuyển bên trong hoặc giữa các tế bào hoặc mô. (b): vận chuyển vào hoặc ra khỏi tế bào.

Protein cũng có thể dự trữ các chất khác ngoài thành phần amino acid (N, C, H, O và S), ví dụ ferritin là protein tìm thấy trong mô động vật kết hợp với Fe. Một phân tử ferritin (460 kDa) gắn với 4.500 nguyên tử Fe (chiếm 35% khối lượng). Protein có vai trò

giữ lại kim loại Fe cần thiết cho sự tổng hợp những protein có chứa Fe quan trọng như hemoglobin.

2.5. Protein vận động và co rút

Một số protein mang lại cho tế bào khả năng vận động, tế bào phân chia và co cơ. Các protein này có đặc điểm: chúng ở dạng sợi hoặc dạng polymer hóa để tạo sợi, chẳng hạn actin và myosin. Tubulin là thành phần cơ bản của thoi vô sắc (sợi xuất hiện khi phân chia các nhiễm sắc thể về các cực).

2.6. Protein cấu trúc

Có chức năng tạo độ chắc và bảo vệ tế bào và mô. Chẳng hạn: α-keratin là protein không tan, cấu tạo nên tóc, sừng và móng. Collagen là protein hình sợi có trong xương. Ở động vật collagen chiếm 1/3 protein tổng số. Fibroin (β-keratin) là thành phần cơ bản của kén tằm.

Một chức năng phổ biến khác của protein là cấu tạo nên màng sinh học.

2.7. Protein bảo vệ

Trong việc giải độc các kim loại nặng, phytochelatin có một ý nghĩa quan trọng, đây là những polypeptide đơn giản có nguồn gốc từ glutation và có công thức chung như sau:

$(\gamma\text{-glutamyl-cysteinyl})_n\text{-glycine}$

Do có nhiều nhóm SH nên chúng có khả năng kết hợp chặt với các kim loại nặng, làm cho những kim loại nặng này không thể gây rối loạn trao đổi chất. Sự tổng hợp phytochelatin được kích thích bởi những kim loại nặng như Cd, Cu, Ag, Bi và Au.

Protein bảo vệ có vai trò quan trọng trong các phản ứng miễn dịch. Động vật có xương sống có một cơ chế phức tạp và phát triển cao để ngăn ngừa những tác nhân vi sinh vật gây bệnh. Chức năng này có liên quan đến đặc tính của chuỗi polypeptide. Khi một protein lạ (có nguồn gốc virus, vi khuẩn hoặc nấm) xâm nhập vào máu hoặc vào mô thì phản ứng tự vệ của cơ thể sẽ xuất hiện rất nhanh. Protein lạ được gọi là kháng nguyên (antigen) chứa một vùng có trật tự xác định các nguyên tử có thể kết hợp với tế bào lympho và kích thích tế bào này sản sinh kháng thể. Những tế bào lympho tồn tại trong hệ thống miễn dịch với số lượng 10^9 và trên bề mặt của nó có những vùng nhận biết nơi mà kháng nguyên sẽ được kết hợp (Hình 1.6). Những vùng nhận biết này rất khác nhau và đặc hiệu cho từng loại kháng nguyên. Trong cơ thể luôn có sẵn một lượng lớn các tế bào lympho khác nhau và chúng có thể tổng hợp rất nhanh các kháng thể đặc hiệu khi

kháng nguyên xuất hiện. Mỗi loại kháng thể có một vị trí kết hợp duy nhất đặc trưng với kháng nguyên. Khả năng bảo vệ của hệ miễn dịch đã làm cho protein lạ của tác nhân gây bệnh trở thành vô hại. Những kháng thể này được gọi là globulin miễn dịch. Chúng chiếm khoảng 20% protein tổng số trong máu.

Một nhóm protein bảo vệ khác là protein làm đông máu thrombin và fibrinogen, ngăn cản sự mất máu của cơ thể khi bị thương.

Cá ở các vùng cực của Trái đất có protein chống đông (antifreeze protein) có tác dụng bảo vệ máu khi nhiệt độ xuống dưới 0^oC.

2.8. Protein lạ/ngoại lai

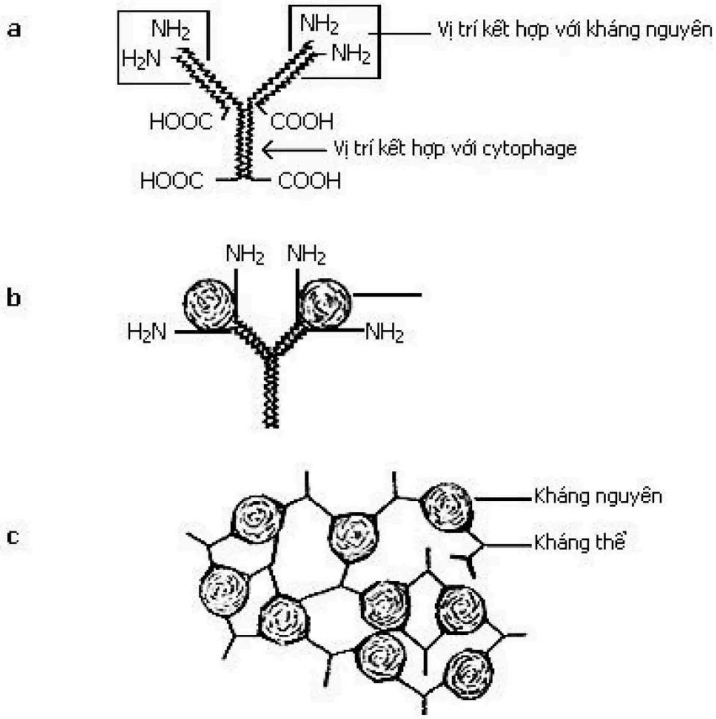
Ví dụ monellin là một loại protein được tìm thấy ở một loại cây ở châu Phi, được coi là chất ngọt nhân tạo cho con người.

Ở một số sinh vật biển như họ Trai tiết ra loại protein keo (glue protein), cho phép nó gắn chặt lên bề mặt.

III. Lipid

Mặc dù không mang hoạt tính sinh học cao như protein nhưng lipid cũng đóng một vai trò đặc biệt trong hệ thống sống. Chúng là nhân tố chính tạo nên các màng sinh học mà nếu thiếu thì mọi hoạt động của protein sẽ không thể phối hợp nhịp nhàng.

Đơn vị cấu trúc của lipid là các acid béo. Mỗi acid béo được cấu tạo từ một mạch carbohydrate (gồm các nguyên tử C và H) gắn với một nhóm carboxyl có tính acid. Các acid béo khác nhau bởi độ dài của chúng, bởi số lượng và vị trí các liên kết đôi. Các acid béo không có liên kết đôi được gọi là các acid béo bão hòa, các acid béo không bão hòa có ít nhất một liên kết đôi.



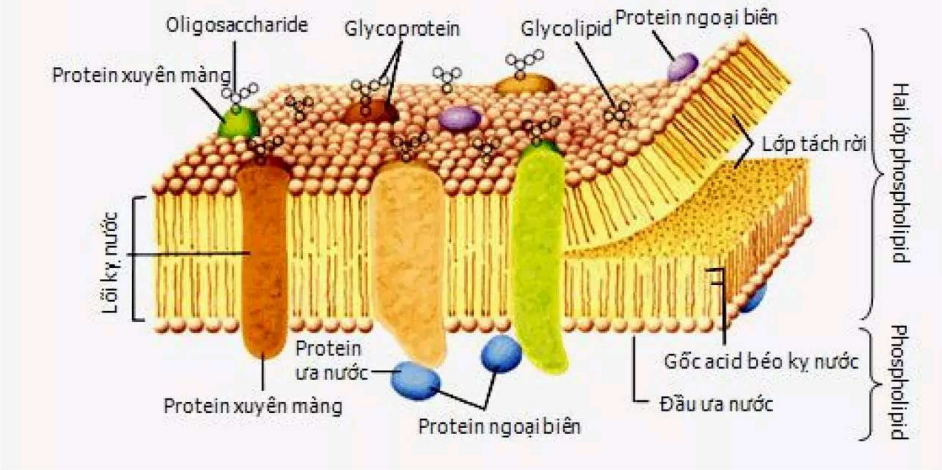
Hình 1.6. Sơ đồ biểu diễn của kháng thể và kháng nguyên. a: kháng thể gồm 4 chuỗi polypeptide. b: kháng thể kết hợp với kháng nguyên. c: kết hợp giữa kháng nguyên và kháng thể.

Màng sinh học có chức năng là giới hạn những vùng trao đổi chất và tham gia vào việc vận chuyển các chất. Màng sinh học cũng có khả năng chuyển đi những tín hiệu. Protein màng cũng có thể là các enzyme. Chức năng này được thể hiện ở màng trong của ty thể và lạp thể. Màng sinh học bao gồm lớp kép lipid với những protein phân bố ở trong đó (Hình 1.7)

Các lipid màng được hình thành từ một chuỗi dài acid béo nối với những nhóm có đặc tính ưa nước cao và được gọi là những phân tử lưỡng cực vì một đầu tương tác với nước, còn đầu kia thì kỵ nước.

Bảng 1.3. Cấu trúc một số acid béo tiêu biểu trong hệ thống sống

Công thức hóa học	Tên thường dùng
Acid béo bão hòa	
$CH_3(CH_2)_{10}$ COOH	Lauric
$CH_3(CH_2)_{14}$ COOH	Palmitic
Acid béo không bão hòa	
$CH_3(CH_2)_7CH{=}CH\,(CH_2)_7$ COOH	Oleic
$CH_3(CH_2)_4$ CH= CH- CH_2 CH= CH $(CH_2)_7$ COOH	Linoleic



Hình 1.7. Sơ đồ biểu diễn một đoạn cắt của màng sinh học

IV. Polysaccharide

Các polysaccharide có nhiều chức năng quan trọng trong tế bào, chúng tham gia vào cấu tạo tế bào và là nguồn dự trữ năng lượng chủ yếu. Các polysaccharide được hình thành từ nhiều monomer, là các đường đơn giản (monosaccharide) nối với nhau bằng liên kết glycoside. Liên kết này được hình thành từ sự kết hợp giữa C_1 của một phân tử đường với nhóm hydroxyl của phân tử kế tiếp.

Nguồn dự trữ tinh bột ở các tế bào động vật là glycogen, trong khi đó ở thực vật là tinh bột. Một polymer khác của glucose là cellulose thì tạo nên thành tế bào thực vật và là hợp chất hữu cơ hiện dịn nhiều nhất trong sinh quyển.

Chúng ta vừa điểm qua riêng rẽ từng thành phần cấu tạo tế bào chính. Trong thực tế, hoạt động của chúng phối hợp mật thiết với nhau. Các nucleic acid trong tế bào thường kết hợp chặt chẽ với các protein tạo thành nucleoprotein. DNA của tế bào

eukaryote thì được bọc bởi những protein đặc hiệu là các histone. Màng tế bào cũng không phải chỉ có phospholipid, chính các protein gắn trong màng đã tạo ra những đặc trưng riêng của màng sinh học. Một điểm cần lưu ý là nếu như cấu trúc và các tính chất hóa lý của nucleic acid, lipid và polysaccharide tương đối đồng nhất thì các protein lại hết sức đa dạng cả về cấu trúc và chức năng. Một phân tử protein thường bao gồm nhiều vùng mang những đặc tính khác nhau: vùng ưa nước hay kỵ nước, vùng gắn một đường, vùng có hoạt tính xúc tác, vùng liên kết với nucleic acid hay với một protein khác. Từ mỗi chức năng của tế bào, từ sự hình thành vật chất mang thông tin di truyền, truyền đạt thông tin di truyền, sự chuyển hóa năng lượng, sự liên lạc giữa các tế bào... đều có sự tham gia của các protein. Điều kỳ diệu của sự sống là toàn bộ các hoạt động vô cùng đa dạng ấy được thực hiện bởi một phân tử duy nhất.

Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. *NXB Giáo dục,* Hà Nội.

2. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

3. Lewin B. 2000. Gene VII. *Oxford University Press,* Oxford, UK.

4. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *WH Freeman and Company*, New York, USA.

5. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Loscik R. 2004. Molecular Biology of the Gene. *The Benjamin Cummings/Cold Spring Habor Laboratory Press*, San Francisco, CA, USA.

6. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company Inc.* New York, USA.

[1]S (Svedberg): đơn vị đo vận tốc lắng. Hệ số lắng của một tiểu đơn vị phụ thuộc không những vào khối lượng của tiểu đơn vị đó mà còn phụ thuộc vào hình dạng và độ rắn của nó, điều này giải thích tại sao sự kết hợp của hai tiểu đơn vị 50S và 30S lại tạo ra một ribosome 70S.

Chương 2

Cấu trúc genome

Genome (hệ gen, bộ gen) là thuật ngữ được dùng với các nghĩa khác nhau như sau:

- Nguyên liệu di truyền của một cơ thể: 1) nhiễm sắc thể trong tế bào vi khuẩn (hoặc một trong mỗi loại nhiễm sắc thể nếu hơn một loại có mặt, ví dụ: các nhiễm sắc thể lớn hoặc bé của *Vibrio cholerae*), 2) DNA hoặc RNA trong một virion, 3) nhiễm sắc thể cùng với mọi plasmid được kết hợp (ví dụ: nhiễm sắc thể và hai plasmid nhỏ trong vi khuẩn *Buchnera*).

- Tất cả các gen (khác nhau) trong tế bào hoặc virion.

- Bộ nhiễm sắc thể đơn bội hoặc genome đơn bội trong tế bào.

Chuỗi genome hoàn chỉnh (nghĩa là trình tự hoàn chỉnh của các nucleotide trong genome) đã được công bố cho một số loài vi khuẩn. Các trình tự khác cũng đã được công bố, ví dụ genome của cây cúc dại (*Arabidopsis thaliana*) và genome người.

Genome chứa toàn bộ thông tin di truyền và các chương trình cần thiết cho cơ thể hoạt động. Ở các sinh vật nhân thật (eukaryote), 99% genome nằm trong nhân tế bào và phần còn lại nằm trong một số cơ quan tử như ty thể và lạp thể. Đa số genome vi khuẩn và phần genome chứa trong các cơ quan tử thường có kích thước nhỏ và ở dạng vòng khép kín. Ngược lại, phần genome trong nhân thường rất lớn và phân bố trên các nhiễm sắc thể dạng thẳng.

Dự án genome là dự án xác định cấu trúc di truyền chính xác của một genome cơ thể sống, nghĩa là trình tự DNA của tất cả các gen của nó. Dự án genome của một số sinh vật mô hình (model organisms) đã được hoàn thành như sau:

- Các genome vi khuẩn. Các trình tự hoàn chỉnh của genome *Escherichia coli* đã được xác định theo phương thức tổ hợp/tập hợp (consortium) của các phòng thí nghiệm. Năm 1995, hai trình tự genome hoàn chỉnh của vi khuẩn *Haemophilus influenzae* và *Mycoplasma genitalium* cũng được hoàn thành. Loài *M. genitalium* có một genome đơn giản (khoảng 580.067 base), do nó dựa vào vật chủ để vận hành nhiều bộ máy trao đổi chất của mình. Loài *H. influenzae* là một vi khuẩn đặc trưng hơn, và có genome khoảng 1.830.121 base với 1.749 gen.

- Chuỗi genome hoàn chỉnh của nấm men *Saccharomyces cerevisiae* đã được hoàn chỉnh trong năm 1996, nhờ một consortium của các phòng thí nghiệm. Genome của chúng dài 12.146.000 base.

- Các dự án genome ở động vật như: chuột, cừu, lợn, giun tròn (*Caenorhabditis elegans*), ruồi giấm (*Drosophila melanogaster*)…, hoặc ở thực vật như: lúa nước, lúa mì, ngô, táo, cúc dại…, mà nổi bật nhất trong số đó là dự án genome người cũng đã được thực hiện.

Ngày 12. 2. 2001 genome người đã được công bố với khoảng 30.000 gen, ít hơn nhiều so với dự kiến trước đây (hàng trăm ngàn gen), và chỉ gấp hai lần giun tròn hoặc ruồi giấm. Người ta đã xác định hệ gen người giống 98% so với tinh tinh và có đến 99% là giống nhau giữa các dân tộc, các cá thể. Do đó, vấn đề hình thành và phát triển nhân

cách, chỉ số thông minh... phải chủ yếu trên cơ sở xã hội và sự rèn luyện của từng người để phát triển tiềm năng sinh học của bản thân.

Trình tự genome của những sinh vật mô hình rất có ý nghĩa trong những nghiên cứu của một chuyên ngành khoa học mới đó là genome học (genomics). Dựa vào đây, các nhà sinh học phân tử có thể phân tích cấu trúc, hoạt động và chức năng của các gen, làm sáng tỏ được vai trò của DNA lặp lại, DNA không chứa mã di truyền, DNA nằm giữa các gen... Điều đặc biệt có ý nghĩa là khi so sánh các genome với nhau, có thể hiểu được hoạt động của genome trong các cơ thể sống, mối quan hệ giữa chúng, sự đa dạng sinh học và mức độ tiến hóa.

Kết quả bước đầu so sánh genome giữa các loài sinh vật với nhau đã cho thấy có ba đặc điểm nổi bật: 1) các gen phân bố trong genome không theo qui luật, 2) kích thước của genome thay đổi không tỷ lệ thuận (tương quan) với tính phức tạp của loài, 3) số lượng nhiễm sắc thể cũng rất khác nhau ngay giữa những loài rất gần nhau.

I. Thành phần và đặc điểm của genome

Genome chứa mọi thông tin di truyền đặc trưng cho từng loài, thậm chí cho từng cá thể trong loài. Genome có thể bao gồm các phân tử DNA hoặc RNA. Đối với sinh vật bậc cao, kích thước genome thay đổi từ 10^9 bp (động vật có vú) đến 10^{11} bp (thực vật). Khác với tế bào tiền nhân (prokaryote), các gen trong genome của eukaryote thường tồn tại nhiều bản sao và thường bị gián đoạn bởi các đoạn mã mù không mang thông tin di truyền (các intron). Vì vậy, một trong những vấn đề đang được quan tâm là cần phải biết số lượng các gen khác nhau có mặt trong genome cũng như số lượng các gen hoạt động trong từng loại mô, từng giai đoạn phát triển và tỷ lệ các gen so với kích thước genome...

1. Genome của cơ quan tử

Hầu hết genome của cơ quan tử, nhưng không phải luôn luôn, có dạng phân tử DNA mạch vòng đơn của một chuỗi duy nhất.

Genome của cơ quan tử mã hóa cho một số, không phải tất cả, các protein được tìm thấy trong cơ quan tử. Do có nhiều cơ quan tử trong một tế bào, cho nên có nhiều genome của cơ quan tử trên một tế bào. Mặc dù bản thân genome của cơ quan tử là duy nhất. Nhưng nó cấu tạo gồm một chuỗi lặp lại[1] liên quan với mọi chuỗi không lặp lại[2] của nhân. Về nguyên tắc, các gen cơ quan tử được phiên mã và dịch mã bởi các cơ quan tử.

1.1. Genome của ty thể

DNA ty thể (mitochondrial DNA-mtDNA) là một genome độc lập, thường là mạch vòng, được định vị trong ty thể.

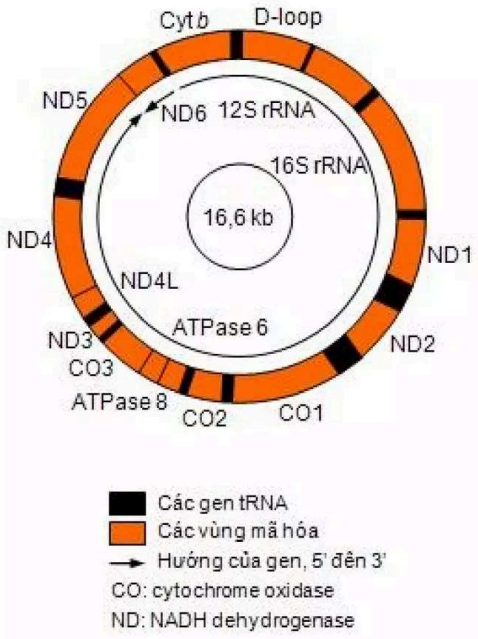
- DNA ty thể của tế bào động vật mã hóa đặc trưng cho 13 protein, 2 rRNA và 22 tRNA.

- DNA ty thể của nấm men *S. cerevisiae* dài hơn mtDNA của tế bào động vật năm lần do sự có mặt của các đoạn intron dài.

Các genome ty thể có kích thước tổng số rất khác nhau, các tế bào động vật có kích thước genome nhỏ (khoảng 16,5 kb ở động vật có vú) (Hình 2.1). Có khoảng một vài trăm ty thể trên một tế bào. Mỗi ty thể có nhiều bản sao DNA. Số lượng tổng số của DNA ty thể so với DNA nhân là rất nhỏ (<1%).

Trong nấm men *S. cerevisiae*, genome ty thể có kích thước khá lớn (khoảng 80 kb) và khác nhau tùy thuộc vào từng chủng. Có khoảng 22 ty thể trên một tế bào, tương ứng khoảng 4 genome trên một cơ quan tử. Ở những tế bào sinh trưởng, tỷ lệ mtDNA có thể cao hơn (khoảng 18%).

Kích thước của genome ty thể ở các loài thực vật là rất khác nhau, tối thiểu khoảng 100 kb. Kích thước lớn của genome đã gây khó khăn cho việc phân lập nguyên vẹn DNA, nhưng bản đồ cắt hạn chế (restriction map) trong một vài loài thực vật đã cho thấy genome ty thể thường là một chuỗi đơn, được cấu tạo như một mạch vòng. Trong mạch vòng này có những chuỗi tương đồng ngắn và sự tái tổ hợp giữa chúng đã sinh ra các phân tử tiểu genome (subgenome) mạch vòng nhỏ hơn, cùng tồn tại với genome "chủ" (master genome) hoàn chỉnh, đã giải thích cho sự phức tạp của các DNA ty thể ở thực vật.



Hình 2.1. DNA ty thể của người. Bao gồm 22 gen tRNA, 2 gen rRNA, và 13 vùng mã hóa protein.

Bảng 2.1 tóm tắt sự phân công của các gen trong một số genome ty thể. Tổng số gen mã hóa protein là khá ít, và không tương quan với kích thước của genome. Ty thể động vật có vú sử dụng các genome 16 kb của chúng để mã hóa cho 13 protein, trong khi đó ty thể nấm men *S. cerevisiae* dùng các genome từ 60-80 kb mã hóa cho khoảng 8 protein. Thực vật với genome ty thể lớn hơn nhiều mã hóa cho nhiều protein hơn. Các intron được tìm thấy trong hầu hết các genome của ty thể, nhưng lại không có trong các genome rất nhỏ của động vật có vú.

Hai rRNA chính luôn được mã hóa bởi genome ty thể. Số lượng các tRNA được mã hóa bởi genome ty thể dao động từ không cho đến đầy đủ (25-26 trong ty thể). Nhiều protein ribosome được mã hóa trong genome ty thể của thực vật và sinh vật nguyên sinh, nhưng chỉ có một ít hoặc không có trong genome của nấm và động vật.

Bảng 2.1. Các genome ty thể có các gen mã hóa cho các protein, rRNA và tRNA

Ty thể mã hóa cho các RNA và protein			
Loài	**Kích thước (kb)**	**Các gen mã hóa protein**	**Các gen mã hóa RNA**
Nấm	19-100	8-14	10-28
Sinh vật nguyên sinh	6-100	3-62	2-29
Thực vật	186-366	27-34	21-30
Động vật	16-17	13	4-24

1.2. Genome của lạp thể

DNA lạp thể (chloroplast DNA-ctDNA) cũng là một DNA genome độc lập, thường là mạch vòng, được tìm thấy trong lạp thể của thực vật.

- Genome của lạp thể rất khác nhau về kích thước, nhưng đủ lớn để mã hóa cho khoảng 50-100 protein cũng như rRNA và tRNA.

- DNA lạp thể dài từ 120-190 kb. Các genome của lạp thể đã được phân tích trình tự cho thấy có khoảng 87-183 gen. Bảng 2.2 mô tả các chức năng được mã hóa bởi genome lạp thể ở cây trồng.

Bảng 2.2. Genome của lạp thể ở các cây trồng mã hóa cho 4 rRNA, 30 tRNA và khoảng 60 protein

Các lạp thể có hơn 100 gen
Các gen
- Mã hóa RNA
16S rRNA
23S rRNA
4,5S rRNA
5S rRNA
tRNA
- Biểu hiện gen
Các r-protein
RNA polymerase
Khác
- Các chức năng của lạp thể
Rubisco và thylakoids
NADH dehydrogenase

Nói chung, các đặc điểm của genome lạp thể tương tự ở ty thể, ngoại trừ lạp thể mang nhiều gen hơn. Genome lạp thể mã hóa cho tất cả các loại rRNA và tRNA cần thiết trong tổng hợp protein, và cho khoảng 50 protein, bao gồm cả RNA polymerase và các protein ribosome.

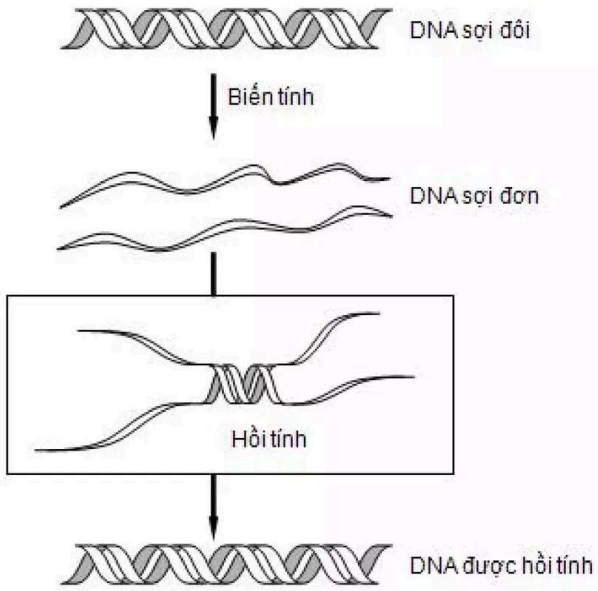
Các intron trong lạp thể được chia thành hai nhóm: 1) những intron ở trên các gen tRNA thường (mặc dù không chắc chắn) được định vị trong vòng anticodon, giống như các intron được tìm thấy trong các gen tRNA của nhân nấm men *S. cerevisiae*; 2) những intron trong các gen mã hóa protein tương tự với các intron của các gen ty thể.

Vai trò của lạp thể là thực hiện quá trình quang hợp. Do đó, nhiều gen của nó mã hóa cho các protein của các phức hợp định vị trong các màng thylakoid. Một vài phức hợp protein của lạp thể giống các phức hợp protein của ty thể: có một số tiểu đơn vị được mã hóa bởi genome của cơ quan tử và một số khác được mã hóa bởi genome của nhân. Nhưng các phức hợp còn lại được mã hóa hoàn toàn bởi genome lạp thể.

2. Động học của phản ứng lai DNA

Bản chất chung của eukaryotic genome được phản ánh qua động học của sự tái liên kết các DNA (DNA reassociation kinetics) bị biến tính. Sự tái liên kết giữa các chuỗi DNA bổ sung xảy ra nhờ bắt cặp base, ngược lại với quá trình biến tính (denaturation) mà nhờ đó chúng được tách rời (Hình 2.2) để thực hiện sự tái bản hoặc phiên mã. Động

học của phản ứng tái liên kết phản ánh sự khác nhau của các chuỗi hiện diện, vì thế phản ứng này có thể được dùng để định lượng các gen và các sản phẩm RNA của chúng.



Hình 2.2. DNA có thể biến tính và hồi tính

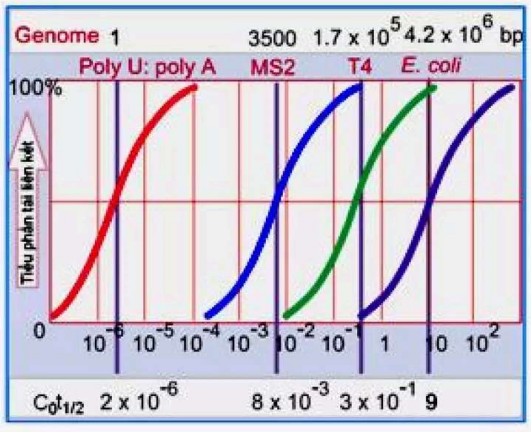
Bảng 2.3 mô tả phản ứng tái liên kết. Sự hồi tính của DNA (renaturation) phụ thuộc vào sự va chạm ngẫu nhiên của các chuỗi bổ sung. Phản ứng của các DNA riêng biệt có thể được mô tả bằng các điều kiện cần thiết cho sự hoàn thành một nửa (half-completion). Đây là tích số của $C_0 \times t_{1/2}$ và được gọi là $C_0t_{1/2}$. Giá trị này tỷ lệ nghịch với hằng số tốc độ. Do $C_0t_{1/2}$ là tích số của nồng độ và thời gian yêu cầu cho một nửa đường, nên một giá trị $C_0t_{1/2}$ lớn hơn dẫn đến một phản ứng chậm hơn.

Bảng 2.3. Một phản ứng tái liên kết của DNA được mô tả bởi $C_0t_{1/2}$

Phản ứng lai phụ thuộc vào C_0t
Tốc độ phản ứng
Phản ứng theo phương trình bậc hai $$\frac{dC}{dt} = -kC^2$$ C là nồng độ của DNA sợi đơn ở thời điểm t. K là hằng số tốc độ tái liên kết.
Tiến độ phản ứng
Lấy tích phân phương trình tốc độ giữa các giới hạn: nồng độ ban đầu của DNA = C_0 ở thời điểm t = 0; nồng độ duy trì sợi đơn = C sau thời gian t $$\frac{C}{C_0} = \frac{1}{1+k.C_0t}$$
Thông số tới hạn là $C_0t_{1/2}$
Khi phản ứng hoàn thành một nửa ở thời điểm t = ½ $$\frac{C}{C_0} = \frac{1}{2} = \frac{1}{1+k.C_0t_{1/2}}$$ Vì thế $C_0t_{1/2} = C_0t_{1/2} = \frac{1}{k}$

Sự hồi tính của DNA thường có dạng đường cong C_0t, đường cong biểu diễn đồ thị phân số của DNA được tái liên kết ($1-C/C_0$) theo log của C_0t. Hình 2.3 trình bày đường cong C_0t của một số genome đơn giản. Các đường cong có dạng tương tự nhau, nhưng giá trị $C_0t_{1/2}$ của mỗi đường là khác nhau.

Các genome trong hình 2.3 đại diện cho các nguồn DNA khác nhau (PolyU:PolyA, thực khuẩn thể MS2, thực khuẩn thể T4 và vi khuẩn *E. coli*). $C_0t_{1/2}$ liên quan trực tiếp với lượng DNA trong genome. Điều này phản ánh tình trạng khi genome trở nên phức tạp hơn, thì sẽ có thêm một số bản sao của một vài chuỗi đặc biệt trong một lượng DNA có trước. Ví dụ: nếu C_0 của DNA là 12 pg, thì nó sẽ chứa khoảng 3.000 bản sao của mỗi trình tự trong genome vi khuẩn.



Hình 2.3. $C_0t_{1/2}$ phụ thuộc vào độ phức tạp của genome. PolyU:PolyA, thực khuẩn thể MS2, thực khuẩn thể T4 và vi khuẩn *E. coli*.

3. Kích thước của genome

Không phải tất cả các đoạn DNA trong genome đều tương ứng với các gen (mã hóa cho protein hoặc một sản phẩm cần thiết cho hoạt động sống của tế bào). Từ những năm 1970, bằng các thí nghiệm gây bão hòa đột biến người ta đã có thể xác định được số gen nằm trên một đoạn nhiễm sắc thể. Ngày nay, nhờ các kỹ thuật phân tích DNA và RNA hiện đại (Southern blot, Northern blot, microarray...) các nhà khoa học có thể xác định số gen hoạt động trong một tế bào. Ví dụ: ở tế bào nấm men *S. cerevisiae* (sinh vật eukaryote bậc thấp) có khoảng 4.000 gen hoạt động, còn tế bào động vật có vú khoảng 10.000-15.000 gen. Như vậy, nếu độ dài trung bình của một gen khoảng 10 kb thì tổng số chiều dài các gen hoạt động trong một tế bào cũng chỉ chiếm 1-2% genome. Hay nói cách khác, chỉ một phần rất nhỏ genome mang thông tin di truyền cần thiết cho hoạt động sống của tế bào. Vậy phần genome còn lại có vai trò gì, và tính phức tạp của loài có liên quan gì với kích thước genome hay không?

Để làm sáng tỏ vấn đề trên, chúng ta cần xem xét kích thước genome của một số loài gần nhau trong bậc thang tiến hóa (có độ phức tạp loài tương tự nhau) cũng như genome của những loài xa nhau (có tính phức tạp khác nhau). Chẳng hạn:

- Genome của người có kích thước khoảng $3,3\times10^9$ bp, trong khi đó genome của những loài lưỡng cư dài khoảng $3,1\times10^9$ bp hoặc thực vật có thể lên đến 10^{11} bp. Như vậy, có phải là các loài lưỡng cư có tính phức tạp tương tự cơ thể chúng ta?

- Hay là ngay trong cùng một loại, chúng ta cũng nhận thấy có sự mâu thuẫn về kích thước genome? Ví dụ: ruồi nhà (*Musca domestica*) có genome khoảng $8,6\times10^8$ bp, lớn gấp sáu lần kích thước genome của ruồi giấm khoảng $1,4\times10^8$bp. Ngoài ra, trong các loài lưỡng cư kích thước genome của chúng cũng thay đổi khá lớn từ 10^9-10^{11} bp. Vì sao ngay trong cùng một loại mà kích thước genome lại biến thiên nhiều như vậy, có phải ruồi nhà có cấu tạo phức tạp hơn nhiều so với ruồi giấm?

Từ những dữ liệu trên, chúng ta có thể nhận định rằng tính phức tạp của loài không liên quan đến kích thước của genome. Tuy nhiên, vai trò của phần genome còn lại (phần không mã hóa) đến nay vẫn chưa được biết nhiều.

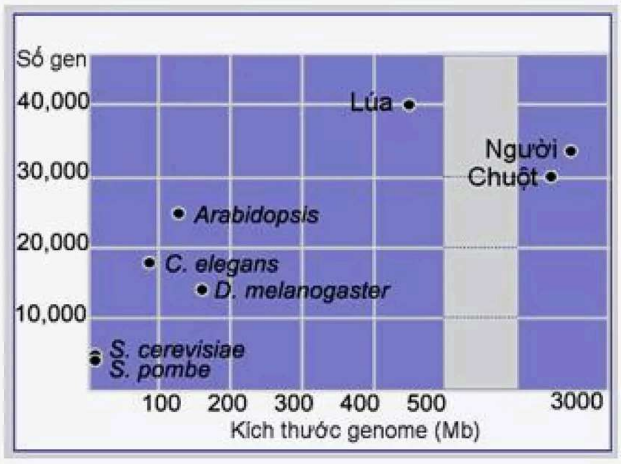
4. Tổng số gen được biết ở một số loài eukaryote

Có 6.000 gen ở nấm men *S. cerevisiae*, 18.500 gen ở giun tròn, 13.600 gen ở ruồi giấm, 25.000 gen ở *Arabidopsis*, và có khả năng 30.000 gen ở chuột và < 30.000 gen ở người.

Như chúng ta đã biết, mối quan hệ giữa kích thước genome và số lượng gen đã không còn nữa. Genome của các sinh vật eukaryote đơn bào có cùng phạm vi kích thước với genome của vi khuẩn lớn nhất. Các eukaryote bậc cao có nhiều gen hơn, nhưng số lượng không tương quan với kích thước genome.

Hình 2.4 cho thấy genome của loài nấm men *S. cerevisiae* dài 13.500 kb và loài nấm men *S. pombe* là 12.500 kb, có khoảng 6.000 gen và 5.000 gen tương ứng. Khung đọc mở trung bình khoảng 1,4 kb, vì thế khoảng 70% genome được chiếm giữ bởi các vùng mã hóa. Sự khác nhau chủ yếu giữa chúng là chỉ 5% gen của *S. cerevisiae* có intron, so với 43% của *S. pombe*.

Genome của giun tròn có khoảng 18.500 gen. Mặc dù genome của ruồi giấm lớn hơn genome của giun tròn, nhưng chúng lại có số gen ít hơn. Đến nay, chúng ta chưa hiểu tại sao ruồi giấm-một cơ thể phức tạp hơn nhiều-chỉ có 70% số gen so với giun tròn. Điều này đã cho thấy không có một mối quan hệ chính xác giữa số gen và tính phức tạp của cơ quan.



Hình 2.4. Số lượng gen của sinh vật eukaryote rất khác nhau. Thay đổi từ 6.000-40.000 nhưng không tương quan với kích thước genome hoặc độ phức tạp của cơ thể.

Cây *Arabidopsis* có kích thước genome trung gian giữa giun tròn và ruồi giấm, nhưng lại có số gen lớn hơn cả hai (25.000). Điều này một lần nữa cho thấy không có một quan hệ rõ ràng, và cũng nhấn mạnh nét đặc biệt của thực vật, là có thể có nhiều gen hơn (do sự nhân đôi của ông bà tổ tiên truyền lại) các tế bào động vật. Đa số genome *Arabidopsis* được tìm thấy trong các đoạn được nhân đôi, gợi ý rằng có một sự nhân đôi xa xưa trong genome (tạo ra một dạng tứ bội). Chỉ 35% các gen của *Arabidopsis* hiện diện như các bản sao đơn.

Genome của lúa lớn hơn *Arabidopsis* khoảng 4 lần, nhưng số gen chỉ lớn hơn khoảng 50%, có khả năng khoảng 40.000 gen. DNA lặp lại chiếm khoảng 42-45% genome. Hơn 80% gen tìm thấy trong *Arabidopsis* được hiện diện trong lúa. Trong số những gen chung này, khoảng 8.000 có trong *Arabidopsis* và lúa nhưng không thấy ở bất kỳ genome động vật hoặc vi khuẩn nào (những gen đã được phân tích trình tự). Có khả năng đây là tập hợp các gen mã hóa cho các chức năng đặc trưng của thực vật, chẳng hạn như quang hợp.

II. Tính phức tạp của genome

Kết quả nghiên cứu động học của các phản ứng lai được tiến hành giữa genomic DNA với cDNA (complementary DNA-DNA bổ sung), giữa DNA với mRNA… cho thấy hầu hết các gen hoạt động đều nằm trong thành phần DNA không lặp lại. Như vậy, thành phần này có ý nghĩa rất quan trọng trong việc đánh giá tính phức tạp của genome. Hay nói cách khác, dựa vào thành phần DNA không lặp lại có thể biết được kích thước genome cũng như mức độ tiến hóa của loài. Nếu như kích thước genome (ở trạng thái đơn bội) được coi là một thông số động học (ký hiệu C), thì giá trị này đặc trưng cho từng loài và không phải luôn luôn tỷ lệ thuận với tính phức tạp của loài. Ngược lại, giá trị C phản ánh các đặc điểm sau:

- Số lượng DNA mã hóa cho các sản phẩm cần thiết đối với hoạt động sống của cơ thể rất nhỏ so với số lượng DNA có trong genome.

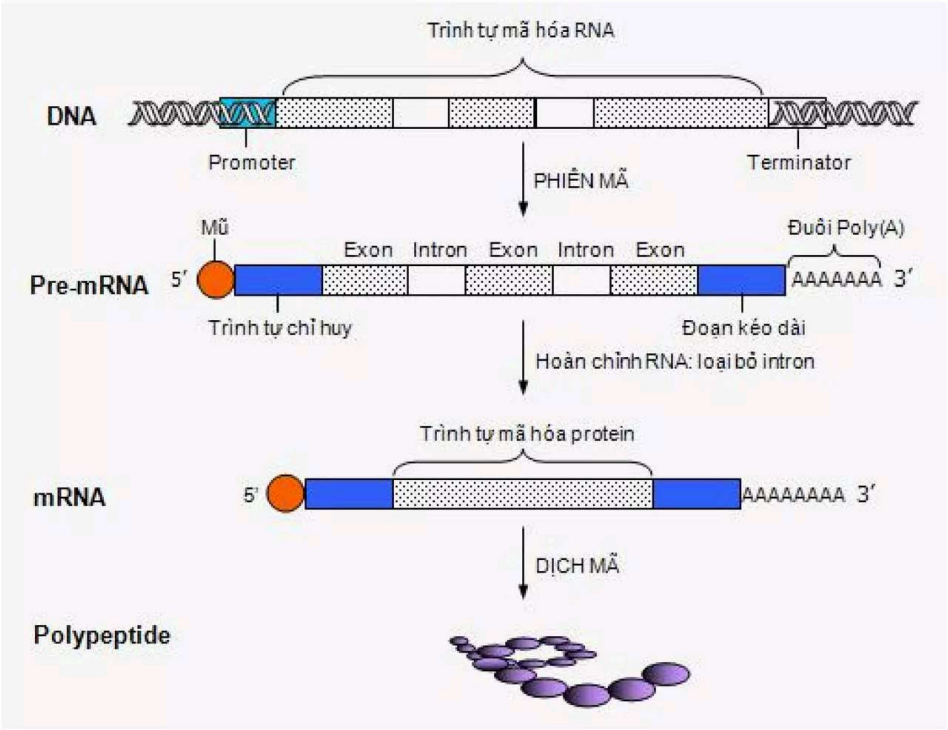
- Có sự biến đổi rất lớn của giá trị C giữa một số loài mà tính phức tạp của chúng không khác nhau nhiều.

Genome vi khuẩn được xem là chỉ chứa các đoạn DNA không lặp lại và các gen thường tồn tại bản sao đơn. Ngược lại, genome của eukaryote thường chứa các gen có hai hoặc nhiều bản sao. Hơn nữa, trình tự nucleotide của các bản sao này có thể không giống nhau hoàn toàn mặc dù sản phẩm protein mà chúng mã hóa có cùng một chức năng. Các bản sao tương đồng của một gen được xếp chung vào một nhóm gọi là một họ gen (gene family). Như vậy, ngoài các gen có một bản sao giống như ở vi khuẩn, genome của eukaryote còn chứa các họ gen. Hầu hết các gen mã hóa cho protein đã được phân lập đều nằm trong các họ gen khác nhau. Các gen trong một họ thường hoạt động theo thời gian và không gian. Điều đó có nghĩa, mỗi thành viên trong họ thường hoạt động ở một thời điểm nhất định trong quá trình hình thành và phát triển cá thể hoặc hoạt động trong

các mô chuyên biệt. Khi một thành viên trong họ bị đột biến (bất hoạt) thì thành viên khác có thể hoạt động thay thế.

Khái niệm về gen được hình thành khi các nhà di truyền học cổ điển nghiên cứu biểu hiện các tính trạng mới do gây đột biến DNA của genome vi khuẩn hay thực khuẩn thể (bacteriophage). Một gen được xem là một đoạn DNA mà bất cứ đột biến nào xảy ra trên đó đều dẫn đến xuất hiện tính trạng mới. Điều này dễ hiểu đối với những genome chỉ chứa các gen có một bản sao (như genome của prokaryote). Tuy nhiên, ở genome của eukaryote, khi có nhiều gen cùng qui định một tính trạng hoặc các gen tương đồng trong một họ có thể hoạt động hỗ trợ thay thế nhau thì đột biến trên một gen không phải lúc nào cũng quan sát được ở mức độ phenotype. Mặt khác, các đoạn DNA tương ứng với trình tự mã hóa (coding sequence, exon) cho một protein thường bị ngắt quãng bởi các đoạn DNA không chứa thông tin di truyền (non-coding sequence, intervening sequence, intron). Các intron được phiên mã cùng với các exon sang phân tử RNA gọi là phân tử tiền thân mRNA (pre-mature mRNA hay pre-mRNA) nhưng sau đó chúng bị loại bỏ và các exon được nối lại với nhau tạo thành phân tử mRNA hoàn chỉnh (mature mRNA hay mRNA) được dùng cho quá trình sinh tổng hợp protein. Quá trình cắt các intron, nối exon không tuân theo một trật tự bắt buộc mà biến hóa đa dạng tạo ra các phân tử mRNA khác nhau từ một phân tử pre-mRNA (Hình 2.5). Bên cạnh mRNA, các phân tử rRNA và tRNA cũng được hình thành từ các phân tử tiền thân chứa intron. Ngoài ra, còn có hiện tượng mã di truyền của gen này nằm xen kẽ với các mã di truyền của gen khác (các gen nằm gối lên nhau-overlapping) hoặc trường hợp dịch chuyển khung đọc ngay trên một đoạn DNA.

Chúng ta đã biết đến chức năng của các phân tử RNA trong hoạt động sống của tế bào. Bên cạnh ba loại RNA đã được nghiên cứu khá kỹ (mRNA, rRNA và tRNA), vai trò của một số loại RNA khác mới được phát hiện vào những năm cuối thế kỷ 20. Chúng kiểm soát sự hoạt động của gen (hiện tượng bất hoạt gen-gene silencing), tham gia phản ứng đọc sửa thông tin di truyền trên phân tử mRNA (hiện tượng RNA editing) hay quyết định tính bền vững của mRNA (các ribonuclease)… Thậm chí có những phân tử pre-mRNA được tổng hợp không phải mã hóa cho protein mà với mục đích phân cắt tạo ra những phân tử RNA có kích thước nhỏ hơn tham gia quá trình kiểm soát hoạt động của các gen khác. Đột biến ở những đoạn DNA mã hóa cho tất cả các loại RNA này thường gắn liền với việc xuất hiện tính trạng mới. Do đó, cần xem xét các đoạn nucleotide đó như các gen mặc dù chúng không mã hóa cho protein.



Hình 2.5. Các gen bị gián đoạn được biểu hiện thông qua RNA tiền thân. Các intron được loại bỏ, trong khi đó các exon được nối lại với nhau. mRNA chỉ có các trình tự của exon được dịch mã thành chuỗi polypeptide.

Ngày nay, theo quan điểm của sinh học phân tử, một gen được xem là một đoạn DNA mã hóa cho một sản phẩm cần thiết đối với hoạt động sống của tế bào. Rõ ràng rằng không phải chỉ có DNA mã hóa cho protein mà cả các DNA mã hóa cho rRNA, tRNA và các loại RNA khác tham gia vào những phương thức kiểm soát hoạt động của genome cũng được xác định là gen.

III. Thay đổi trật tự của các đoạn DNA trong genome-Transposon

1. Sự thay đổi trật tự của các đoạn DNA trong genome

Như đã biết kích thước và cấu trúc genome của các loài rất khác nhau. Nguyên nhân của sự đa dạng này là do: 1) trao đổi chéo giữa các cặp nhiễm sắc thể tương đồng xảy ra trong phân bào giảm nhiễm đã dẫn đến sự đa dạng trong loài; 2) trong genome trật tự các đoạn DNA cũng như cấu trúc các gen được sắp xếp lại, đặc trưng cho từng cá thể. Chính các yếu tố di truyền có khả năng di chuyển giữa các vị trí trong một genome hoặc giữa các genome khác nhau đã góp phần làm đa dạng di truyền giữa các cá thể trong loài.

Các yếu tố di truyền có khả năng di chuyển được xếp vào ba nhóm chính tùy thuộc vào tính độc lập của chúng:

- Nhóm thứ nhất gồm các yếu tố có khả năng di chuyển giữa các vị trí khác nhau trong genome.

- Nhóm thứ hai gồm các yếu tố có khả năng ghép vào và tách ra khỏi genome để tồn tại độc lập trong tế bào (các episome như plasmid F, bacteriophage).

- Nhóm thứ ba chỉ di chuyển dưới sự kiểm soát của tế bào ở những giai đoạn sinh trưởng phát triển nhất định để sắp xếp khởi động một số gen đặc biệt (hình thành cassette hoạt động ở nấm men *S. cerevisiae*, ký sinh trùng đơn bào *Trypanosome*).

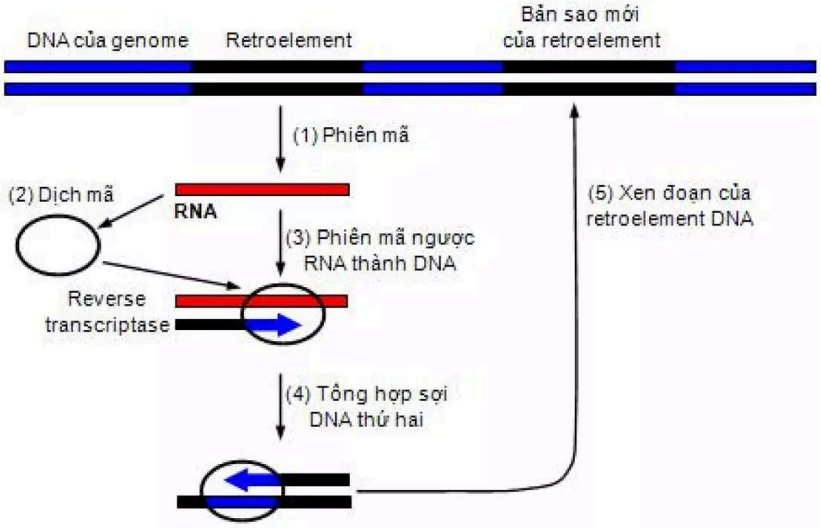
Sự di chuyển của các yếu tố ở nhóm thứ hai và thứ ba liên quan tới tái tổ hợp tương đồng hoặc tái tổ hợp ở các vị trí đặc hiệu. Mặc dù không có khả năng tồn tại độc lập bên ngoài genome, nhóm thứ nhất tự kiểm soát sự di chuyển của chúng trong genome và không đòi hỏi sự tương đồng giữa chúng với vị trí ghép vào. Do đó, có thể nói chúng di chuyển một cách tự do trong genome và được gọi tên chung là các yếu tố di chuyển (transposable elements, transposons). Trong khi thực khuẩn thể λ được xem là episome, thì thực khuẩn thể *Mu* và một số retrovirus ở eukaryote được xem là transposable elements.

2. Các transposon

Các yếu tố di chuyển có thể được chia làm hai loại dựa vào cách thức di chuyển của chúng:

- Loại thứ nhất được gọi là retroelement, chúng phải trải qua hình thức trung gian RNA trong quá trình di chuyển (RNA genome được sao chép nhờ reverse transcriptase tạo ra cDNA để ghép vào vị trí mới) (Hình 2.6).

- Loại thứ hai là các đoạn DNA hoặc bị tách ra hoặc được tái bản tạo thêm bản sao để ghép vào vị trí mới. Thông thường, loại thứ hai này được gọi là transposon.



Hình 2.6. Retroelement

Sự tồn tại của các transposon (còn gọi là gen nhảy) là nét đặc trưng của tế bào thực vật. Ở sinh vật eukaryote, các transposon còn được gọi là yếu tố kiểm soát (controlling elements). Transposon có thể di chuyển từ nhiễm sắc thể này sang nhiễm sắc thể khác hoặc ở các vị trí khác nhau trên cùng một nhiễm sắc thể. Chúng có thể "nhảy" vào giữa dãy mã của một gen đang hoạt động làm cho gen này bất hoạt hoặc ngược lại. Ở cây ngô, người ta đã phát hiện 10 nhóm transposon và một số trong chúng đã được giải mã. Tất cả chúng đều có kích thước khoảng 4.200 nucleotide và khác nhau chủ yếu ở đoạn mã tận cùng của gen.

Khi di chuyển, các transposon gây ra việc sắp xếp và tổ chức lại genome của từng cá thể như tạo các đoạn DNA mới ở vị trí chúng ghép vào và tách ra. Chúng có thể di chuyển tới vị trí bất kỳ và hoàn toàn không cần một mối quan hệ nào giữa hai vị trí mới và cũ. Khi tách ra transposon có thể mang theo các đoạn DNA bên cạnh, gây ra sự mất đoạn tại vị trí cũ. Ngược lại, khi ghép vào vị trí mới, chúng gây ra hiện tượng thêm đoạn hoặc chuyển đoạn ở vị trí mới. Do đó, transposon giống như các vector chuyên chở DNA từ nơi này sang nơi khác trong một genome hoặc từ genome này sang genome khác.

Ngoài ra, trao đổi chéo giữa các transposon tương đồng ở hai vị trí khác nhau trên một hoặc trên hai nhiễm sắc thể cũng tạo ra những biến đổi tương tự. Những biến đổi đó dẫn đến việc sắp xếp lại genome, tạo ra tính đa dạng giữa chúng và tính đặc thù riêng của từng cá thể. Đặc biệt, sự thay đổi vị trí của các transposon còn có thể ảnh hưởng đến hoạt động của các gen phân bố xung quanh ngay khi chúng không làm thay đổi trật tự nucleotide ở những gen này. Tần số ghép của các transposon vào genome khoảng 10^{-5}-10^{-7} sau mỗi thế hệ. Ngược lại, tần số tách ra khoảng 10^{-6} đến 10^{-10}.

Các transposon có thể chia làm hai nhóm dựa vào khả năng di chuyển độc lập hay phải phụ thuộc vào sự có mặt của transposon khác:

- Nhóm thứ nhất gồm các đoạn DNA có khả năng di chuyển độc lập. Chúng chứa gen mã hóa cho các protein điều khiển quá trình đó, ví dụ như enzyme nhận biết hai đầu transposon để tách chúng ra khỏi vị trí cũ và ghép vào vị trí mới. Do đó, quá trình này được thực hiện hoàn toàn độc lập. Nhờ khả năng này, chúng tạo ra các đột biến không bền vững.

- Nhóm thứ hai gồm các transposon không có khả năng tự hoạt động, tức là chúng không có khả năng di chuyển do không mang đủ thông tin di truyền mã hóa cho các enzyme cần thiết. Vì vậy, các transposon loại này tạo ra những đột biến gen một cách tự phát nhưng là đột biến bền vững. Việc di chuyển của transposon ở nhóm này phụ thuộc vào sự có mặt của transposon có khả năng hoạt động độc lập cùng nhóm. Hai transposon có thể xếp vào cùng nhóm khi chúng có cấu trúc tương đồng nhau, đặc biệt là các đoạn oligonucleotide phân bố ở hai đầu transposon. Đây là vị trí để enzyme nhận biết và cắt nối transposon ở vị trí đi và đến.

Các transposon đơn giản nhất ở vi khuẩn được gọi là IS (insertion sequences). Chúng có thể nằm trên chromosome hoặc trên các plasmid. Transposon vi khuẩn không giữ một chức năng nào trong tế bào. Trình tự nucleotide ở một đầu IS thường lặp lại nhưng ngược chiều so với đầu kia (inverted repeat). Ví dụ như GGTAT-X_n-ATACC. Do đó, khi sợi đôi IS tách thành hai sợi đơn thì mỗi sợi này có khả năng hình thành liên kết bổ sung tại hai đầu của IS tạo cấu trúc thân-quai/cán-thòng lọng (stem-loop).

Transposon thường mã hóa cho các enzyme transposase làm nhiệm vụ nhận biết chuỗi nucleotide lặp lại ngược chiều (inverted repeat) để cắt transposon và di chuyển. Khi một IS được ghép vào vị trí bất kỳ của genome thì một đoạn ngắn DNA tại vị trí này được nhân đôi, tức là hai đầu của IS được chặn bởi các nucleotide giống nhau, sắp xếp theo cùng một chiều. Vì vậy, đoạn ngắn DNA này được gọi là "lặp lại xuôi chiều" (direct repeat). Chiều dài của chúng thường khoảng 9 bp (Hình 2.7). Dựa vào sự có mặt của các đoạn cùng chiều và ngược chiều có thể xác định được vị trí transposon ghép vào hoặc chuyển đi.

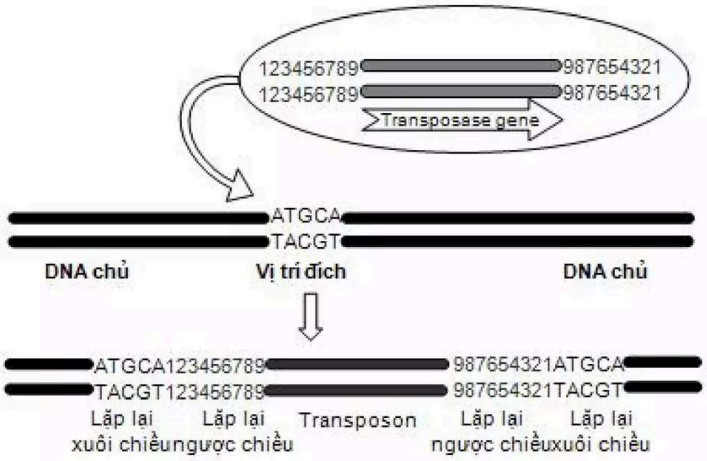
Ngoài các IS, ở vi khuẩn còn có các đoạn DNA có khả năng di chuyển với kích thước dài hơn, gọi là T_n. Các T_n thường phân bố trên plasmid (phân tử DNA dạng vòng, kích thước thường không lớn) và có khả năng chèn vào bất kỳ vị trí nào trong genome. Chúng thường mang thông tin di truyền mã hóa cho các protein kháng kháng sinh. Giữa IS và T_n có mối quan hệ về trình tự các nucleotide. Các T_n thường được giới hạn ở hai đầu bởi một loại IS nào đó.

Quá trình di chuyển của một transposon từ vị trí cũ (vị trí cho, donor) sang vị trí mới (vị trí nhận, recipient) xảy ra theo hai cơ chế khác nhau:

- **Cơ chế sao y bản chính (transposon có mặt ở cả hai vị trí).** Theo cơ chế này, phiên bản sau khi được sao chép từ vị trí cho sẽ được ghép vào vị trí nhận. Như vậy, mỗi

lần di chuyển thì số lượng bản sao được tăng lên. Quá trình này liên quan đến hai loại enzyme: transposase (tác động vào hai đầu bản gốc transposon) và resolvase (tác động lên bản sao).

- **Cơ chế tách ra khỏi vị trí cũ di chuyển đến vị trí mới.** Theo cơ chế này, một transposon có thể tách ra khỏi vị trí cũ và ghép vào vị trí mới. Như vậy, số lượng transposon không thay đổi. Kiểu di chuyển này chỉ cần enzyme transposase. Khi transposon chuyển đi, vị trí cũ bị gãy và nó được nối lại nhờ cơ chế sửa chữa DNA trong tế bào.



			Chiều dài tổng số	Chọn lọc đích
IS1	9 bp	23 bp	768 bp	ngẫu nhiên
IS2	5 bp	41 bp	1327 bp	các điểm nóng
IS4	11-13 bp	18 bp	1428 bp	$AAAN_{20}TTT$
IS5	4 bp	16 bp	1195 bp	các điểm nóng
IS10R	9 bp	22 bp	1329 bp	NGCTNAGCN
IS50R	9 bp	9 bp	1531 bp	các điểm nóng
IS903	9 bp	18 bp	1057 bp	ngẫu nhiên

Hình 2.7. Một transposon có đoạn lặp lại ngược chiều gồm 9 nucleotide (123456789) gắn vào vị trí có 5 nucleotide (ATGCA). Đoạn ngắn ATGCA có mặt ở cả hai đầu của IS và có trình tự nucleotide sắp xếp theo cùng một chiều.

Khi transposon ghép vào vị trí mới, một đoạn nucleotide ngắn ở đó được sao thành hai bản, mỗi bản nằm ở một đầu của transposon (direct repeat). Tại vị trí nhận, mỗi sợi đơn DNA bị cắt lệch nhau vài nucleotide. Transposon nối vào các đầu cắt, tạo ra hai khoảng trống (gaps). Chúng được sửa chữa theo nguyên tắc tạo cặp bổ sung. Do đó, đoạn nucleotide nằm giữa hai vết cắt được sao chép thành hai bản, mỗi bản ở một đầu và trình tự sắp xếp các nucleotide giống nhau. Vì vậy, chúng được gọi là lặp lại xuôi chiều.

IV. Tương tác của T-DNA với genome thực vật

Sự di chuyển DNA từ genome vi khuẩn sang genome thực vật được nghiên cứu khá kỹ đối với tương tác giữa *Argobacterium tumefaciens* hoặc *A. rhizogenes* với hầu hết các cây hai lá mầm. Hiện tượng di chuyển DNA này gây những biến đổi về mặt di truyền, biểu hiện ở việc xuất hiện các khối u trên thân cây (*A. tumefaciens*) hoặc mọc rất nhiều rễ tơ (*A. rhizogenes*) tại nơi bị nhiễm vi khuẩn.

A. tumefaciens và *A. rhizogenes* là hai loài vi khuẩn gây bệnh ở thực vật. Tuy nhiên, sau đó bệnh được duy trì lại không phụ thuộc sự tồn tại của vi khuẩn. Đó là do một số gen của vi khuẩn đã được chuyển vào genome cây chủ và hoạt động gây bệnh.

Các gen vi khuẩn có khả năng di chuyển và hoạt động trong tế bào thực vật nằm trên Ti-plasmid (tumor inducing plasmid) của *A. tumefaciens* hoặc trên Ri-plasmid (hairy-root inducing plasmid) của *A. rhizogenes*. Cũng như các khối u động vật, các tế bào thực vật có DNA vi khuẩn ghép vào genome bị chuyển sang trạng thái mới, ở đó sự phát triển và biệt hóa của chúng hoàn toàn khác với các tế bào bình thường. Đó là do hoạt động của các gen vi khuẩn trong genome của thực vật. Bình thường, những gen này có mặt trong genome vi khuẩn nhưng chúng chỉ được hoạt động (mở) sau khi hợp nhất vào genome thực vật và chịu sự kiểm soát của tế bào cây chủ. Quá trình này có tính chất đặc hiệu vật chủ, tức là một loại vi khuẩn chỉ có khả năng gây khối u trên một số loại cây chủ này mà không tương tác được với các loại cây khác.

Việc tạo khối u hay thực hiện quá trình chuyển gen từ vi khuẩn sang genome thực vật dẫn đến biến đổi trạng thái sinh lý của tế bào thực vật đòi hỏi các điều kiện sau:

- Phải có hoạt động của các gen trên ba vùng *chvA, chvB*, *pscA* nằm trên nhiễm sắc thể của vi khuẩn để khởi động việc bám dính vi khuẩn vào thân cây.

- Ti-plasmid hoặc Ri-plasmid phải mang vùng *vir* (nằm ngoài vùng T-DNA). Vùng này mang các gen cần thiết cho việc tách và vận chuyển T-DNA sang tế bào thực vật.

- Các gen trên vùng T-DNA của Ti-plasmid hoặc Ri-plasmid được ghép vào genome thực vật gây biến đổi trạng thái các tế bào này.

1. Ti-plasmid và Ri-plasmid

Plasmid là các vòng DNA tự sinh sản độc lập ở bên ngoài nhân. Ở vi khuẩn và động-thực vật, plasmid liên quan tới yếu tố giới tính của tế bào, đến khả năng chống chịu các loại kháng sinh... Đặc điểm quan trọng của plasmid là chúng có thể liên kết vào nhiễm sắc thể nhưng cũng có thể tồn tại bên ngoài nhiễm sắc thể một cách độc lập.

Các plasmid của *Agrobacterium* mang các vùng T-DNA, *vir* (virulence region), gen chuyển hóa opine, và vùng *ori* khởi đầu sao chép trong tế bào vật chủ.

Sự khác nhau giữa Ti-plasmid và Ri-plasmid ở chỗ, vùng T-DNA của Ti-plasmid chứa các gen tổng hợp auxin, cytokinin và opine. Trong khi đó, vùng T-DNA của Ri-plasmid chỉ mang gen tổng hợp auxin và opine. Đây là điểm khác nhau gây ra hiệu quả

tác động khác nhau lên thực vật. Cơ chế gây bệnh của các *Agrobacterium* là sau khi xâm nhiễm vào tế bào, chúng gắn đoạn T-DNA vào bộ máy di truyền của tế bào thực vật, làm rối loạn các chất sinh trưởng nội sinh, tạo ra khối u (trường hợp *A. tumefaciens*) hoặc rễ tơ (trường hợp *A. rhizogenes*). Khả năng chuyển các gen này đã được khai thác để chuyển gen ngoại lai vào bộ máy di truyền của tế bào thực vật theo ý muốn.

2. T-DNA

Vùng T-DNA được nghiên cứu rất kỹ. Đó là một đoạn DNA có kích thước 25 kb trong đó chứa gen mã hóa cho sinh tổng hợp auxin, cytokinin và opine (trường hợp Ti-plasmid) hoặc auxin và opine (trường hợp Ri-plasmid). Trong plasmid, vị trí của T-DNA được giới hạn bằng biên phải (right border-RB) và biên trái (left border-LB), mỗi biên có chiều dài 25 bp.

3. Vùng vir

Trong các vùng DNA của Ti-plasmid và Ri-plasmid, ngoài T-DNA, được nghiên cứu nhiều hơn cả là vùng DNA phụ trách khả năng lây nhiễm còn gọi là vùng *vir* (virulence). Sản phẩm hoạt động của các gen nằm trong vùng *vir* dưới tác động kích thích của các hợp chất phenol tiết ra từ vết thương là một loạt các protein đặc hiệu như *vir*A, *vir*E2, *vir*B, *vir*D, *vir*D2, *vir*C1... Các protein này nhận biết các vết thương ở các cây chủ thích hợp (hầu hết là cây hai lá mầm), kích thích sản sinh ra các đoạn T-DNA, bao bọc che chở các đoạn DNA này và giúp chúng tiếp cận với genome của cây chủ một cách an toàn.

4. Quá trình chuyển T-DNA vào tế bào thực vật

Quá trình cắt đoạn T-DNA ra khỏi plasmid và vận chuyển nó vào tế bào thực vật trước hết phụ thuộc vào sản phẩm của các gen *vir*. Vi khuẩn xâm nhiễm tại bất kỳ vị trí tổn thương nào trên thân cây. Cây có vết thương do sự hỏng ngẫu nhiên của màng tế bào thực vật hoặc do vi khuẩn tiết ra hỗn hợp những chất được mã hóa bởi các gen *vir*. Hoạt động của các gen này được hoạt hóa bởi hợp chất phenolic của cây (ví dụ như acetosyringone, catechol, các dẫn xuất của chalcone...). Ngoài ra, các monosaccharide như glucose, arabinose có mặt trong môi trường cũng làm tăng khả năng gây cảm ứng nhóm gen *vir*.

Protein VirA đóng vai trò quan trọng trong việc quyết định loại cây chủ bị nhiễm bởi *Agrobacterium*. Trong thực tế, *Agrobacterium* không có khả năng xâm nhập vào cây một lá mầm, có thể protein VirA không nhận biết các tín hiệu do cây một lá mầm tiết ra. Nói chung, *Agrobacterium* chỉ gây hại ở cây hai lá mầm, vì vậy người ta cho rằng chúng chỉ có thể đưa T-DNA vào hệ gen các cây hai lá mầm. Tuy nhiên, gần đây nhiều tác giả đã chứng minh khi nhiễm vi khuẩn, các cây một lá mầm cũng có thể sản xuất opine và vì thế có thể khai thác khả năng biến nạp gen của *Agrobacterium* vào cây một lá mầm. Hiện nay, người ta cũng đã thành công trong việc chuyển gen vào một số cây một lá mầm như lúa, mía... qua trung gian *A. tumefaciens.*

Khi cây nhiễm *A. tumefaciens*, do T-DNA hợp nhất vào trong genome của cây chủ bắt đầu hoạt động và sản xuất ra auxin, cytokinin và opine, toàn bộ sinh trưởng của cây bị rối loạn, các tế bào phân chia vô tổ chức và tạo ra các khối u. Opine được vi khuẩn sử dụng như một loại "thức ăn" nhờ gen chuyển hóa opine trên Ti-plasmid. Cơ chế lây nhiễm của *A. rhizogenes* đối với cây hai lá mầm cũng tương tự, nhưng trong vùng T-DNA của *A. rhizogenes* chỉ có gen sản sinh ra auxin, vì thế sự thay đổi hình thái chính của thực vật là chúng tạo ra rất nhiều rễ tơ khi bị nhiễm bệnh.

Để gắn T-DNA vào tế bào thực vật, đầu tiên vi khuẩn *A. tumefaciens* phải tiếp xúc với thành tế bào thực vật bị tổn thương. Quá trình này được thực hiện nhờ các gen *chvA* và *chvB*. Gen *chvB* mã hóa một protein liên quan đến hình thành β-1,2 glucan mạch vòng, trong khi đó gen *chvA* xác định một protein vận chuyển, định vị ở màng trong của tế bào vi khuẩn. Protein vận chuyển giúp vận chuyển β-1,2 glucan vào khoảng giữa thành tế bào và màng sinh chất. β-1,2 glucan giữ vai trò quan trọng để vi khuẩn *Agrobacterium* tiếp xúc với thành tế bào thực vật. Nếu không có sự tiếp xúc này, sẽ không có sự dẫn truyền T-DNA.

Các sản phẩm protein của vùng *vir* có tác dụng cho việc dẫn truyền T-DNA từ vi khuẩn vào tế bào thực vật. Các loại protein đó rất cần thiết cho quá trình cắt T-DNA khỏi plasmid, cảm ứng thay đổi màng tế bào thực vật mà chúng tiếp xúc, tham gia di chuyển phần T-DNA qua màng vi khuẩn tới tế bào chất của tế bào thực vật, vận chuyển tới nhân rồi cuối cùng xâm nhập vào genome của cây chủ.

Thực chất, chỉ riêng T-DNA của plasmid được chuyển vào genome tế bào thực vật, mà không còn phần nào khác. Quá trình dẫn truyền chỉ do sản phẩm của các gen *vir* và gen *chv* quyết định mà không liên quan đến các gen khác trên T-DNA. Tuy nhiên, chuỗi DNA 25 bp (RB và LB của T-DNA) có vai trò là vị trí cảm ứng cho các sản phẩm của tổ hợp các gen vùng *vir*, đặc biệt là protein từ gen *vir*E mang chúng dẫn truyền vào tế bào thực vật. Chúng hoạt động như các tín hiệu nhận biết và khởi động quá trình dẫn truyền. Trước hết gen *vir*A trong tổ hợp gen vùng *vir* được phosphoryl hóa nhờ tác động của các hợp chất phenol như acetosyringone giải phóng ra từ các tế bào thực vật tổn thương. Sản phẩm của quá trình này lại tiếp tục phosphoryl hóa gen *vir*G. Sản phẩm của gen *vir*G liên tiếp làm hoạt hóa toàn bộ các gen *vir* còn lại, mà hai gen cuối cùng được hoạt hóa là gen *vir*B và *vir*E. Trước đó, khi gen *vir*D được hoạt hóa, sản phẩm của nó cảm ứng nhận biết RB và LB của T-DNA và làm đứt phần T-DNA ra khỏi DNA của plasmid thành các sợi đơn. Đồng thời, quá trình phosphoryl hóa này cũng làm thay đổi thẩm suất màng tế bào thực vật, màng tế bào bị mềm ra và bị thủng. Các sợi đơn T-DNA được gắn vào protein do gen *vir*E tổng hợp và dịch chuyển về phía màng tế bào vi khuẩn. Ngay sau đó, sợi T-DNA được trượt từ vi khuẩn vào tế bào thực vật. Cầu nối chính là sự tiếp hợp (conjugation) giữa hai tế bào do cảm ứng sản phẩm gen *vir*B mà thành. Khi T-DNA đã được chuyển giao vào tế bào thực vật, chúng nhanh chóng hợp nhất vào genome tế bào thực vật được ổn định và di truyền như các gen bình thường khác.

V. Sắp xếp và khuếch đại các gen trong genome

1. Sắp xếp lại các gen

Nói chung, genome có cấu trúc và tổ chức bền vững. DNA của genome thường không bị biến đổi bởi sự phát triển vô tính, nhưng thỉnh thoảng các trình tự của chúng cũng có thể bị chuyển chỗ trong gen, bị cải biến, khuếch đại hoặc thậm chí biến mất, như là một trường hợp tự nhiên.

Trao đổi chéo trong phân bào giảm nhiễm là một trong những nguyên nhân gây ra biến đổi của genome. Tuy nhiên, điều này xảy ra chủ yếu trong tế bào sinh dục mà không có trong các tế bào soma. Việc sắp xếp lại genome sẽ dẫn đến những thay đổi sau:

- Tạo ra các gen mới cần thiết cho sự biểu hiện trong các trường hợp đặc biệt.

- Sự tái sắp xếp có thể đáp ứng cho việc mở hoặc đóng gen. Đây cũng chính là cơ chế của sự điều hòa biểu hiện gen.

Một ví dụ điển hình là hiện tượng sắp xếp lại các gen trong genome của nấm men *S. cerevisiae* và của ký sinh trùng *Trypanosome* châu Phi khi gây chứng ngủ li bì ở vật chủ.

1.1. Chuyển đổi dạng giao phối của nấm men

Nấm men có thể tồn tại ở cả hai dạng đơn bội hoặc lưỡng bội. Các dạng lưỡng bội là dị hợp tử ở trường hợp locus kiểu giao phối, và các tế bào đơn bội có thể hoặc là MATa hoặc MATα. Việc chuyển đổi trạng thái được thông qua giao phối, kết hợp các bào tử đơn bội thành lưỡng bội và qua việc tái tạo giao tử mới. Tuy nhiên, giao phối chỉ xảy ra giữa hai loại tế bào đơn bội a và α. Các tế bào đơn bội cùng loại không thể kết hợp với nhau tạo thành tế bào lưỡng bội.

MAT (mating type locus) là vị trí hoạt động hay cassette hoạt động: một vùng đặc biệt của nhiễm sắc thể số 3 chứa các gen qui định dạng giao phối (a và α). Các gen này còn tồn tại ở hai vị trí khác trong genome. Tuy nhiên, ở đó chúng đều bị bất hoạt. Hai vị trí này được gọi là hai vị trí tĩnh (hay cassette tĩnh HML và HMR). Mỗi vị trí tĩnh chỉ mang các gen qui định cho một dạng giao phối. Khi các gen được sao chép từ một vị trí tĩnh vào vị trí hoạt động thì mRNA mới được tổng hợp từ các gen đó. Như vậy, quá trình sao chép quyết định dạng giao phối của nấm. Bản gốc luôn được bảo tồn ở vị trí tĩnh và bản thứ hai xuất hiện ở vị trí hoạt động.

Nếu vị trí tĩnh có mang đột biến thì chúng sẽ được sao chép vào vị trí hoạt động. Tuy nhiên, nếu xảy ra đột biến ở cassette MAT thì tính trạng mới xuất hiện không bền. Một khi dạng giao phối chuyển đổi thì các gen cũ ở vị trí MAT bị thay thế bởi bản sao của vị trí tĩnh khác. Lúc đó, đột biến bị loại đi và tính trạng mới sẽ biến mất.

So sánh giữa các dạng cùng sợi nấm (homothallic) và khác sợi nấm (heterothallic) người ta nhận thấy dạng heterothallic có gen *HO* hoạt động và có thể chuyển đổi tự động giữa các kiểu giao phối, và vì thế một bào tử đơn có thể làm tăng quần thể tự phối, trong khi dạng homothallic không có gen *HO* và duy trì cùng một kiểu giao phối trong suốt chu trình sinh trưởng đơn bội.

Phân tích di truyền cho thấy các gen sau đây cần cho việc chuyển đổi kiểu giao phối:

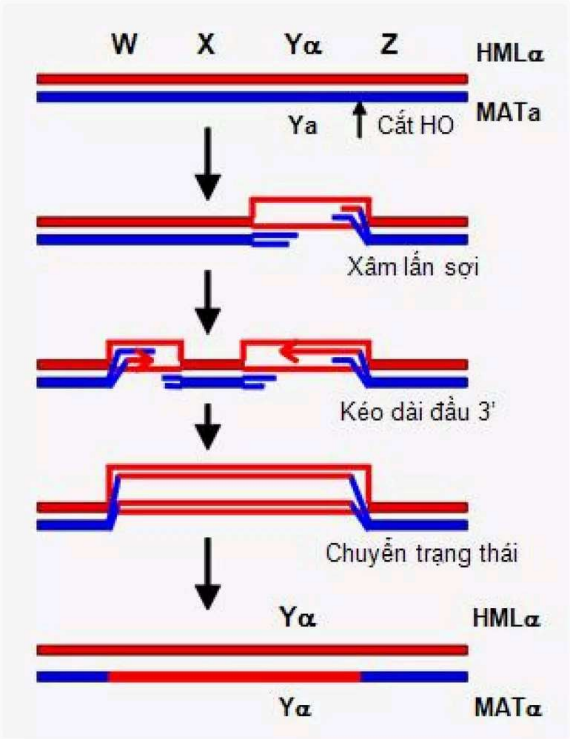
- *MAT*
- *HO,* mã hóa cho enzyme endonuclease
- *HML*ꞅ (cho MATa chuyển thành MATα)
- *HMR*a (cho MATα chuyển thành MATa)

Trình tự các nucleotide ở vị trí tĩnh và vị trí hoạt động chỉ khác nhau một đoạn ngắn ký hiệu là Ya và Yα (Hình 2.8). Enzyme HO-endonuclease nhận biết vị trí đặc hiệu tại ranh giới phân cách giữa Z và Y và của cassette hoạt động MAT và cắt cả hai sợi DNA tại đó. Điều đặc biệt là enzyme này không cắt DNA khi chúng hiện diện ở cassette tĩnh.

Sau khi đoạn Y của vùng MAT bị phân hủy hết, đoạn Y của một trong hai cassette tĩnh được dùng làm khuôn mẫu để sao chép vào vị trí bị phân hủy (Hình 2.8). Nếu đột biến xuất hiện ở vùng MAT (đoạn Y), tính trạng mới chỉ biểu hiện tạm thời. Một khi dạng giao phối chuyển đổi, các gen bình thường được sao chép vào vùng MAT và đột biến bị loại đi.

1.2. Chuyển đổi gen ở *Trypanosome*

Trypanosome có khả năng lẩn tránh được hệ thống miễn dịch của vật chủ thường bằng cách thay đổi kháng nguyên bề mặt (surface antigen) của chúng. Mỗi loại kháng nguyên được tổng hợp nhờ hoạt động của một gen tương ứng tại vị trí hoạt động. Gen này có thể bị thay thế bởi một gen mã hóa cho loại kháng nguyên khác nằm ở một vị trí tĩnh nào đó trong genome. Mỗi vị trí tĩnh chứa một gen ở trạng thái không hoạt động. Gen này chỉ được mở khi chuyển đến vị trí hoạt động. Trong genome của *Trypanosome*, có rất nhiều vị trí tĩnh nhưng chỉ có một vị trí hoạt động.



Hình 2.8. Quá trình chuyển đổi dạng giao phối từ a sang α nhờ trao đổi gen giữa vùng MATa và HMLα trên nhiễm sắc thể số 3

Bình thường, *Trypanosome* sẽ trải qua một số lần biến đổi hình thái khi được truyền từ ruồi châu Phi sang vật chủ. Bề mặt của tế bào *Trypanosome* được bao bọc một lớp đơn gồm 5×10^6 phân tử của một loại glycoprotein (VSG-variable surface glycoprotein). Đây chính là kháng nguyên bề mặt của *Trypanosome* khi chúng xâm nhập vào vật chủ. Điều đáng chú ý là chúng có khả năng thay đổi kháng nguyên bề mặt, do đó tránh được phản ứng miễn dịch của tế bào vật chủ. Quá trình thay thế kháng nguyên bề mặt phụ thuộc vào sự chuyển đổi các gen mã hóa cho chúng xảy ra ở một vị trí đặc biệt trong genome (vị trí hoạt động). Chuyển đổi gen mã hóa kháng nguyên bề mặt nhằm mục đích hoạt hóa gen mã hóa kháng nguyên bề mặt mới thay thế cho kháng nguyên tồn tại trước đó. Khi một gen đang hoạt động bị thay thế bởi một gen khác sẽ tương ứng với việc xuất hiện kháng nguyên mới và loại bỏ kháng nguyên cũ.

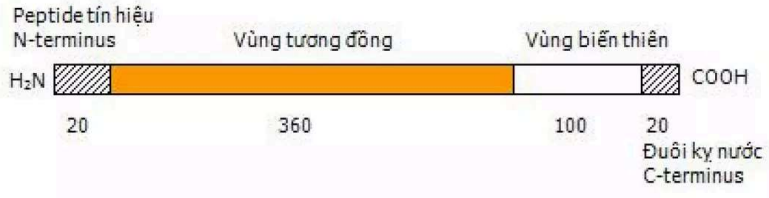
- Cấu trúc của một VSG ở *Trypanosome*

Cấu trúc chung của một VSG được mô tả trên hình 2.9 và 2.10. Một VSG vừa được tổng hợp dài khoảng 500 amino acid gồm tín hiệu N-terminus, tiếp theo là đoạn peptide quyết định tính kháng nguyên; đoạn peptide bảo thủ giữa các VSG và đuôi kỵ nước. Phân tử này được tổng hợp dưới dạng protein tiền thân (pre-protein). Do đó, chúng phải trải qua biến đổi ở hai đầu NH_2 và COOH để trở thành dạng protein hoàn chỉnh (mature form). Dạng này được đính vào màng tế bào ở đầu COOH.

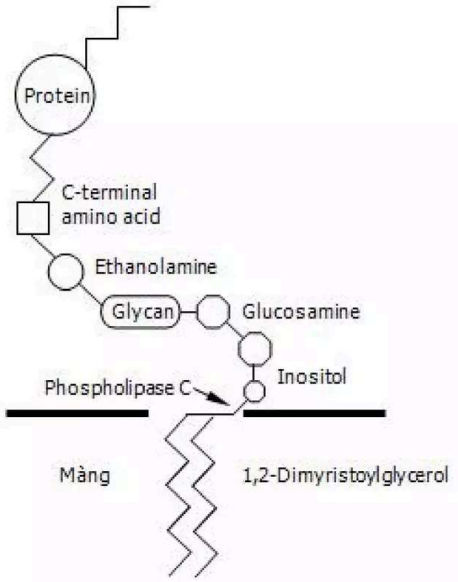
Một loại *Trypanosome* có thể tạo ra ít nhất khoảng 100 VSG từ khi nhiễm cho đến khi gây chết vật chủ. Số gen mã hóa cho VSG có thể nhiều hơn 1.000 gen, tất cả các gen này đều nằm trong genome. Tuy nhiên, tại một thời điểm bất kỳ chỉ có một gen hoạt động tổng hợp nên một loại VSG. Do đó, sự thay đổi kháng nguyên tương ứng với sự thay đổi hoạt động của gen. Khi một gen mới được mở, gen hoạt động trước nó phải bị ức chế hoàn toàn. Lúc đó, một kháng nguyên mới sẽ thay thế kháng nguyên tồn tại trước nó.

Hình 2.9 cho thấy chuỗi polypeptide chứa khoảng 500 amino acid. N-terminus chứa một peptide tín hiệu cho sự vận chuyển qua ER (lưới nội sinh chất, endoplasmic reticulum) và tới màng plasma, được tách ra khỏi protein hoàn chỉnh. Vùng biến thiên là khác nhau ở mỗi VSG, điều đó cho thấy các VSG có ít hoặc không có tính đồng nhất. Hướng tới phần C-terminus, chuỗi polypeptide được bảo toàn tốt hơn và phần này được gọi là vùng tương đồng. Đuôi kỵ nước chứa một tín hiệu nhận biết để gắn với mỏ neo glycolipid (glycolipid anchor) (Hình 2.10). Khi mỏ neo được gắn thì 20 amino acid cuối cùng sẽ được tách ra.

Glycoprotein biến đổi bề mặt (VSG) được gắn với màng thông qua một mỏ neo glycolipid chứa ethanolamine, một cấu trúc glycan mang một số gốc mannose (mannose moiety), một glucosamine và một phosphoinositol liên kết với 1,2-dimyristoyglycerol có gai trong màng plasma. Gốc glycolipid là yếu tố quyết định phản ứng lai (cross-reacting) (CRD) được nhận biết bằng các kháng thể phản ứng với tất cả dạng biến đổi của VSG, nhưng chỉ khi VSG được phóng thích khỏi màng. Màng liên kết với VSG không được nhận biết bởi các kháng thể anti-CRD. Sự phóng thích VSG được xúc tác bởi hoạt tính của trypanosome-specific phospholipase C được giả định là hiện diện ở mặt bên trong (inner face) của màng plasma. Người ta không biết rằng enzyme có thể cắt liên kết phosphoester trên mặt khác của màng như thế nào.



Hình 2.9. Sơ đồ minh họa chuỗi protein của một VSG đặc trưng



Hình 2.10. Cấu trúc của mỏ neo glycolipid của VSG

- Hoạt động của gen VSG

Gen mã hóa cho một VSG được gọi là bản gen gốc (basic copy gene). Các gen này được phân thành hai nhóm tùy thuộc vào vị trí của chúng trên nhiễm sắc thể.

+ Các gen nằm ở telomere (khoảng 5-15 kb)>200 gen.

+ Các gen nằm cách telomere hơn 50 kb.

Tương tự như ở nấm men, các gen mã hóa cho VSG nằm rải rác trong genome và ở trạng thái không hoạt động. Một gen bất kỳ được hoạt hóa chỉ khi nó được sao chép vào vị trí hoạt động (expression site) trong khi nguyên bản của nó vẫn được bảo tồn ở vị trí tĩnh. Bản sao của bản gen gốc vào vị trí hoạt động được gọi là ELC (bản sao hoạt động-expression linked copy). Vị trí hoạt động nằm ở gần telomere. Như vậy, việc chọn một bản gen gốc để sao chép tạo bản sao hoạt động sẽ phụ thuộc vào vị trí của bản gen gốc ở telomere hoặc nằm phía trong telomere. Có hai giả thiết để một gen trở nên hoạt hóa như sau:

- Vị trí hoạt động không thay đổi mà chỉ có các ELC thay thế cho nhau. Bản sao của một bản gen gốc sẽ thay thế cho bản sao của một gen khác ở tại vị trí đó, điều này xảy ra tương tự như các gen ở cassette tĩnh được sao chép vào cassette hoạt động trong trường hợp với nấm men.

- Vị trí hoạt động thay đổi, khi cần tổng hợp một VSG mới, gen ở vị trí hoạt động cũ bị ngừng và gen ở một vị trí khác (gần telomere) được khởi động.

Một số phân tử mRNA tương ứng với các VSG khác nhau được phân lập và được xác định trình tự nucleotide (thông qua cDNA). Điều ngạc nhiên là phần oligonucleotide ở đầu 3' của mọi gen VSG đều khác với phần 3' của các mRNA được phiên mã từ các gen đó. Mặt khác, các gen này không có phần 5' giống như các mRNA. Như vậy, các mRNA không được tổng hợp hoàn toàn trên khuôn mẫu các gen này. Phần 3' của các mRNA tương ứng với phần 3' của vị trí hoạt động ELC, trong khi phần 5' (gồm 35 nucleotides) được tổng hợp từ những đoạn DNA khác và được gắn vào mRNA (hiện tượng trans-splicing).

2. Khuếch đại các gen

Số lượng bản sao của một số gen cũng có thể được tăng lên tạm thời trong quá trình phát triển các tế bào soma. Việc tăng số lượng của một gen đặc biệt nào đó phụ thuộc vào từng điều kiện cụ thể của tế bào và xảy ra không phổ biến. Các bản sao có thể nằm tập trung thành một nhóm gồm bản sao này nối tiếp bản sao khác hoặc có thể tồn tại như những đoạn DNA có khả năng tái bản độc lập. Chẳng hạn:

- Sự nhân bản của gen mã hóa cho rRNA ở trứng ếch. Trứng ếch có đường kính khoảng 2-3 mm, dự trữ rất nhiều rRNA. Chúng được phiên mã từ rất nhiều gen rDNA. Các gen này được nhân lên (khoảng 2.000 lần) theo cơ chế "vòng tròn quay" (xem chương 4) trong quá trình phát triển và tồn tại dưới dạng các vòng tròn khép kín.

- Khi nuôi cấy các tế bào động vật có vú trong môi trường đặc biệt, DNA tại một số vị trí trong genome được nhân lên. Ví dụ: nuôi cấy các tế bào ung thư trong môi trường chứa độc tố methotrexate. Chất này ức chế hoạt tính của enzyme dihydrofolate reductase (DHFR) giữ vai trò trong tổng hợp các nucleotide của DNA. Các tế bào ung thư nuôi cấy trong môi trường có chất độc này phát triển thành các quần lạc tế bào kháng lại độc tố. Khi nồng độ chất độc tăng dần, nồng độ DHFR cũng tăng theo, có thể đạt tới 1.000 lần lớn hơn mức bình thường. Nồng độ enzyme tăng do số lượng các gen mã hóa cho chúng tăng. Cơ chế chính xác của hiện tượng này chưa rõ ràng, nhưng có thể xảy ra theo hai cách:

- Trao đổi chéo không cân bằng giữa hai nhiễm sắc tử (chromatid) của nhiễm sắc thể dẫn đến một số tế bào không có gen *dhfr* và một số khác có hai bản sao của gen này. Trong môi trường có độc tố, trao đổi chéo không cân bằng được lặp đi lặp lại và các tế bào chứa nhiều gen *dhfr* vẫn phát triển tốt trong môi trường này.

- Các đoạn DNA (100-1.000 kb) chứa 2-4 gen *dhfr* (~31 kb/gen) được sao chép từ nhiễm sắc thể bình thường tạo ra các nhiễm sắc thể rất nhỏ, không có tâm động. Các nhiễm sắc thể nhỏ này ghép vào các nhiễm sắc thể bình thường khác. Quá trình này lặp đi lặp lại và qua một số lần phân bào nguyên nhiễm, tế bào nào mang số lượng lớn các gen *dhfr* càng có điều kiện phát triển thuận lợi trong môi trường có chứa độc tố.

3. Biến nạp gen

Một phương thức tăng khả năng di truyền là ứng dụng các tương tác vật chủ cộng sinh-ký sinh, trong đó DNA lạ được chuyển vào tế bào vật chủ từ một vi khuẩn. Cơ chế này tương tự với sự tiếp hợp của vi khuẩn. Sự biểu hiện của DNA vi khuẩn trong vật chủ mới của nó sẽ làm thay đổi kiểu hình của tế bào. Ví dụ điển hình là vi khuẩn *A. tumefaciens* cảm ứng tạo khối u ở tế bào thực vật bị chúng xâm nhiễm.

Khi DNA lạ được đưa vào tế bào eukaryote, nó có thể tồn tại ngoài nhiễm sắc thể hoặc được hợp nhất trong genome. Nếu xảy ra theo trường hợp thứ hai, genome sẽ mang những đột biến di truyền và nhiều khi DNA lạ vẫn tiếp tục hoạt động làm xuất hiện các tính trạng mới. Việc đưa các gen lạ vào tế bào soma hoặc tế bào sinh dục mà vẫn duy trì hoạt động của những gen đó được gọi là chuyển nhiễm (transfection). Cá thể biểu hiện tính trạng mới nhờ hoạt động của gen lạ đưa vào tế bào sinh dục được gọi là cá thể chuyển gen. Quá trình này có thể làm thay đổi sự ổn định của genome. DNA sau khi được tiêm vào trong các tế bào trứng động vật có thể được hợp nhất trong genome và truyền cho thế hệ sau như một thành phần di truyền bình thường. Khả năng đưa một gen chức năng đặc trưng có thể trở thành một kỹ thuật y học sử dụng cho việc chữa bệnh di truyền.

Tài liệu tham khảo/đọc thêm

1. Võ Thị Thương Lan. 2002. Sinh học phân tử. *NXB Đại học Quốc gia Hà Nội*, Hà Nội.

2. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. Garland Publishing, Inc. New York, USA.

3. Cantor CR and Smith CL. 1999. Genomics. *John Wiley & Sons, Inc.* New York, USA.

4. Dale JW and Von Schanzt M. 2002. From Gene to Genome. *John Wiley & Sons, Ltd.* West Sussex, UK.

5. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

6. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *WH Freeman and Company*, New York, USA.

7. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Loscik R. 2004. Molecular Biology of the Gene. *The Benjamin Cummings/Cold Spring Habor Laboratory Press*, San Francisco, CA, USA.

8. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company Inc.* New York, USA.

[1]DNA lặp lại mô tả các chuỗi hiện diện hơn một bản sao trong mỗi genome.
[2]DNA không lặp lại chứa các chuỗi duy nhất: chỉ có một bản sao trong genome đơn bội.

Chương 3

Cấu trúc và chức năng của gen

I. Định nghĩa gen

Chúng ta có thể điểm qua những mốc chính trong lịch sử nghiên cứu về gen như sau:

Mendel (1865) là người đầu tiên đưa ra khái niệm **nhân tố di truyền**. Johansen (1909) đã đề xuất thuật ngữ **gen** (từ **genos**, nghĩa là sản sinh, nguồn gốc) để chỉ nhân tố di truyền xác định một tính trạng nào đó. Sau đó, Morgan trong những năm 1920 đã cụ thể hóa khái niệm về gen, khẳng định nó nằm trên nhiễm sắc thể và chiếm một locus nhất định, gen là đơn vị chức năng xác định một tính trạng.

Vào những năm 1940, Beadle và Tatum đã chứng minh gen kiểm tra các phản ứng hóa sinh và nêu giả thuyết **một gen-một enzyme**. Tuy nhiên, trường hợp hemoglobin là một protein nhưng lại gồm hai chuỗi polypeptide do hai gen xác định, do đó giả thuyết trên buộc phải điều chỉnh lại là **một gen-một polypeptide**.

Vào những năm 1950, DNA (deoxyribonucleic acid) được chứng minh là vật chất di truyền. Mô hình **cấu trúc DNA** của Watson và Crick được đưa ra và **lý thuyết trung tâm** (central dogma) ra đời. Gen được xem là một đoạn DNA trên nhiễm sắc thể mã hóa cho một polypeptide hay RNA.

Cuối những năm 1970, việc phát hiện ra gen gián đoạn ở sinh vật eukaryote cho thấy có những đoạn DNA không mã hóa cho các amino acid trên phân tử protein. Vì thế, khái niệm về gen lại được chỉnh lý một lần nữa: Gen là một đoạn DNA đảm bảo cho việc tạo ra một polypeptide, nó bao gồm cả phần phía trước là vùng 5' không dịch mã (5' untranslation) hay còn gọi là vùng ngược hướng (upstream) và phía sau là vùng 3' không dịch mã (3' untranslation) hay còn gọi là vùng cùng hướng (downstream) của vùng mã hóa cho protein, và bao gồm cả những đoạn không mã hóa (intron) xen giữa các đoạn mã hóa (exon).

Hiện nay, có thể định nghĩa gen một cách tổng quát như sau: Gen là đơn vị chức năng cơ sở của bộ máy di truyền chiếm một locus nhất định trên nhiễm sắc thể và xác định một tính trạng nhất định. Các gen là những đoạn vật chất di truyền mã hóa cho những sản phẩm riêng lẻ như các mRNA được sử dụng trực tiếp cho tổng hợp các enzyme, các protein cấu trúc hay các chuỗi polypeptide để gắn lại tạo ra protein có hoạt tính sinh học. Ngoài ra, gen còn mã hóa cho các tRNA, rRNA và snRNA...

Bảng 3.1. Tóm tắt lịch sử nghiên cứu về di truyền học

Mốc thời gian	Năm	Các sự kiện chính
1850	1865	Gen là các nhân tố hạt
	1871	Khám phá ra nucleic acid
1900	1903	Nhiễm sắc thể là các đơn vị di truyền
	1910	Gen nằm trên nhiễm sắc thể
	1913	Nhiễm sắc thể là các dãy sắp xếp mạch thẳng của gen
	1927	Đột biến là những thay đổi vật lý của gen
	1931	Sự tái tổ hợp xuất hiện bởi hiện tượng vắt chéo
	1944	DNA là vật liệu di truyền
	1945	Gen mã hóa cho protein
1950	1951	Trình tự protein đầu tiên
	1953	DNA có dạng xoắn kép
	1958	DNA tái bản theo phương thức bán bảo thủ
	1961	Mã di truyền là bộ ba
	1977	Các gen của sinh vật eukaryote bị gián đoạn
	1977	DNA có thể được phân tích trình tự
	1995	Genome của vi khuẩn được phân tích trình tự
2000	2001	Genome người được phân tích trình tự

II. Lý thuyết trung tâm

1. Sự xác định di truyền cấu trúc bậc một của protein

Cấu trúc không gian của chuỗi polypeptide được xác định bởi trình tự sắp xếp của các amino acid tức cấu trúc bậc một. Như vậy, mặc dù có nhiều mức độ cấu trúc không gian khác nhau, nhưng cấu trúc bậc một tức trình tự sắp xếp các amino acid chi phối toàn bộ các mức độ cấu trúc khác. Việc xác định di truyền phân tử protein ở trạng thái tự nhiên có đầy đủ hoạt tính sinh học chỉ quy tụ lại chủ yếu ở xác định cấu trúc bậc một là đủ.

2. Các enzyme mất hoạt tính do đột biến

Nhiều nghiên cứu cho thấy, việc mất hoạt tính enzyme nhiều khi không phải do vắng mặt của enzyme, mà chỉ do các biến đổi trên phân tử (modification). Có trường hợp đột biến dẫn đến những thay đổi tinh vi, enzyme vẫn có hoạt tính nhưng sẽ biểu hiện khác nếu thay đổi điều kiện. Chẳng hạn:

Ở nấm mốc *Neurospora crassa*, enzyme tyrosinase do gen *T* xác định, xúc tác cho phản ứng chuyển hóa tyrosine thành dihydroxyphenylalanine. Alelle T^+ của dòng hoang dại sản xuất tyrosinase có hoạt tính ở nhiệt độ bình thường và cả ở 60°C. Một đột biến T^S sản xuất tyrosine có hoạt tính ở nhiệt độ bình thường, nhưng lại mất hoạt tính ở 60°C.

Như vậy, trong đa số trường hợp, đột biến của một gen không làm biến mất enzyme mà chỉ biến đổi cấu trúc dẫn đến thay đổi hoạt tính. Các đột biến của cùng một gen có thể gây ra những biến đổi khác nhau trên enzyme. Các hiện tượng đó chứng tỏ rằng cấu trúc của enzyme chịu sự kiểm soát trực tiếp của gen.

3. Bản chất các biến đổi di truyền của protein

Bản chất đó chính là quan hệ một gen-một polypeptide.

Như đã nêu trên, người ta khám phá ở người có những gen tạo ra hemoglobin (Hb) khi biến dị sẽ tạo ra những hemoglobin bất thường do sai hỏng ở các chuỗi polypeptide α hoặc β (Bảng 3.2 và 3.3) và gây ra các bệnh di truyền.

Bảng 3.2. Các loại hemoglobin ở chuỗi polypeptide α

Thứ tự amino acid	1	2	16	57	58	68	116	141
Tên amino acid	Val	Leu	Lys	Glu	His	Asp	Glu	Arg
Loại hemoglobin *Hb I*			Asp					
Hb Nocfolk				Asp				
Hb M Boston					Tyr			
Hb B Philadelphia						Lys		
Hb O Indonesia							Lys	

Ở trường hợp này, thứ tự các amino acid có thể bị sai lệch. Từ những dữ liệu trên ta thấy các dạng đột biến tạo một Hb bất thường đó là thay một amino acid này bằng một amino acid khác.

Bảng 3.3. Các loại hemoglobin ở chuỗi polypeptide β

Thứ tự amino acid	1	2	3	6	26	67	121	146
Tên amino acid	Val	His	Leu	Glu	Glu	Val	Glu	His
Loại hemoglobin *Hb S*				Val				
Hb C				Lys				
Hb E					Lys			
Hb M Minvauki						Glu		
Hb O Á Rập							Lys	

Qua hai chuỗi polypeptide α và β chúng ta thấy có một số dạng hemoglobin bất thường ở người. Trên mỗi chuỗi, chỉ trình bày những amino acid đã bị thay đổi ở dạng đột biến. Số thứ tự chỉ vị trí của amino acid trong chuỗi polypeptide. Mỗi hemoglobin bất thường có thể được đặt cho một chữ cùng tên (nếu có) của địa phương được tìm thấy.

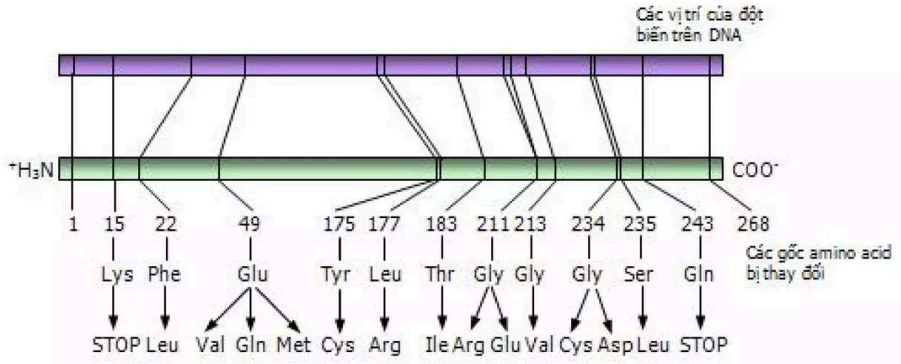
Đột biến được biểu hiện bởi sự thay thế vị trí của một amino acid này bằng một amino acid khác.

4. Sự tương quan đồng tuyến tính gen-polypeptide

4.1. Đột biến tryptophan synthetase-sự đồng tuyến tính giữa gen và chuỗi polypeptide

Nghiên cứu trên enzyme tryptophan synthetase xúc tác cho phản ứng tổng hợp tryptophan của *E. coli* người ta nhận thấy có nhiều đột biến xảy ra trên cùng một gen mã hóa cho tryptophan synthetase.

Thực hiện tái tổ hợp trong gen (nguyên tắc là gen ở các vị trí càng xa nhau trên nhiễm sắc thể càng dễ tái tổ hợp), người ta đã nhận được các dạng biến dị có tính chất khác nhau, và tính được khoảng cách tương đối giữa những điểm khác nhau của đột biến đã được xác định. Vị trí biến dị trên thể nhiễm sắc tương ứng với vị trí của amino acid trên chuỗi polypeptide. Như vậy, có thể cho rằng có sự đồng tuyến tính giữa gen và chuỗi polypeptide (Hình 3.1).



Hình 3.1. Tương quan đồng tuyến tính giữa gen và enzyme tryptophan synthetase của *E. coli* thông qua các vị trí đột biến và các gốc amino acid bị thay đổi

Nhiều dạng đột biến của tryptophan synthethase đã được tạo ra. Bằng cơ chế tái tổ hợp, những khoảng cách tương đối giữa những điểm khác nhau của đột biến đã được xác định. Sản phẩm protein của mỗi dạng đột biến đã được phân tích, và những thay đổi các amino acid khác cũng được xác định. Người ta đã tìm thấy mối tương quan hoàn toàn giữa những khoảng cách của các đột biến được tìm thấy trên gen với khoảng cách của amino acid bị thay đổi trong phân tử protein.

4.2. Đột biến

4.2.1. Khái niệm

Một gen (DNA) có 4 loại base và một phân tử protein có 20 loại amino acid[1], nhưng giữa chúng có mối tương quan như thế nào. Đầu tiên, người ta cho rằng một base qui định một amino acid, nhưng những tính toán cho thấy không hợp lý. Vì chỉ có 4 base trong DNA và 20 amino acid trong protein, cho nên mỗi codon phải chứa ít nhất 3 base. Hai base cũng không thể làm thành một codon bởi vì chỉ có $4^2 = 16$ cặp hợp lý của 4 base. Nhưng 3 base thì có thể bởi vì sẽ có $4^3 = 64$ bộ ba hợp lý. Vì số lượng bộ ba hợp lý lớn hơn 20, cho nên sẽ có trường hợp một vài codon chỉ định cùng một amino acid. Ví dụ: UCU, UCC, UCA, UCG, AGU và AGC đều cùng mã hóa cho serine.

Từ đó, người ta đưa ra khái niệm mã di truyền (tín hiệu di truyền). Mã di truyền cho phép đọc thứ tự trên DNA để biết thứ tự trên chuỗi polypeptide. Mã di truyền không mơ hồ, có nghĩa với một trình tự chẳng hạn ATA ta biết nó ghi mã cho một amino acid gì, và cũng thấy rằng có nhiều mã di truyền xác định cho một amino acid (Bảng 3.4).

4.2.2. Đột biến điểm

Là đột biến chỉ tác động một vị trí, nói rõ hơn đó là một base. Khi thay đổi một base trên DNA sẽ tạo ra sự thay đổi một amino acid (Hình 3.2).

Đột biến dĩ nhiên xảy ra trên DNA và được sao lại trên mRNA trong phiên mã, rồi trên protein trong dịch mã.

Bảng 3.4. Mã di truyền chung

Vị trí thứ nhất	Vị trí thứ hai				Vị trí thứ ba
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
U	Leu (L)	Ser (S)	STOP	STOP	A
U	Leu (L)	Ser (S)	STOP	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
C	Leu (L)	Pro (P)	His (H)	Arg (R)	C
C	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
C	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
A	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
A	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
G	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
G	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

Chú thích

Những đơn vị mã (codon) được đọc theo chiều 5'→3'.

STOP: codon kết thúc (còn gọi là vô nghĩa).

Đột biến điểm có các dạng sau:

- Đột biến sai nghĩa. Thay đổi một amino acid trong protein, có thể dẫn đến một trong ba kết quả sau:

+ Không hậu quả nào cả, vì amino acid không nằm trong vị trí hoạt động hoặc không có vai trò trong cấu trúc enzyme.

+ Có biến đổi nhẹ ở chuỗi polypeptide sẽ tạo ra tính mẫn cảm yếu với nhiệt, làm giảm sự ổn định chuỗi polypeptide.

+ Mất hẳn hoạt tính enzyme nếu đúng ngay vị trí hoạt động của enzyme đó.

- Đột biến vô nghĩa. Thay đổi một base. Nếu đó là một codon vô nghĩa sẽ làm ngừng kéo dài (tổng hợp) chuỗi polypeptide ở vị trí amino acid này. Tức là nếu codon này nằm ở đầu sẽ không có chuỗi polypeptide hoạt động.

- Đột biến acridine hoặc đột biến dịch khung. Đột biến này do chất acridine màu da cam tạo ra (hoặc còn gọi là đột biến dịch khung, frameshift, do thêm vào hoặc bớt đi một base) (Hình 3.2 E và D). Như vậy, một đột biến trên khung đọc khi thêm vào (C) hoặc mất đi (A) thường sẽ dẫn đến xuất hiện một **codon stop** làm ngừng chuỗi polypeptide và enzyme sẽ không có hoạt tính.

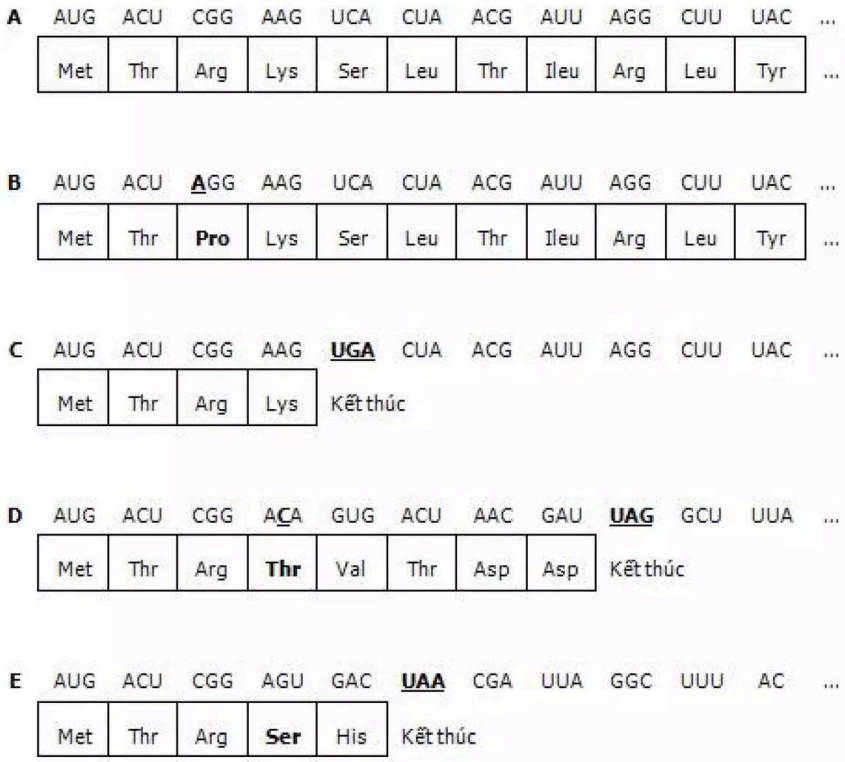
4.2.3. Đột biến kìm hãm

Đến nay, người ta nhận thấy mọi sai lệch trong việc tổng hợp protein nếu có đều xảy ra từ DNA, còn quá trình diễn ra từ RNA đến polypeptide luôn luôn đúng. Nghiên cứu một vài kiểu protein đột biến ta thấy:

- Đột biến sai nghĩa. Làm xuất hiện một bất thường trong trình tự amino acid. Kết quả protein mất hoạt tính. Hoạt tính này có thể được phục hồi, hoặc do một đột biến ngược để cho lại protein cấu trúc ban đầu.

- Đột biến vô nghĩa. Làm mất đi một phần chuỗi polypeptide, phần còn lại không có hoạt tính, và hoạt tính này có thể có lại được nhờ đột biến trong một codon đã bị đột biến.

Thông thường, những gen kìm hãm đột biến vô nghĩa không nằm ở gần vị trí của đột biến ấy. Đó là những gen làm biến đổi hệ thống dịch mã khi tổng hợp protein.



Hình 3.2. Các dạng đột biến điểm

A: trình tự các codon và các amino acid tương ứng ở dạng tự nhiên.

B: thay đổi một base (C thành A) làm thay đổi một amino acid (Arg thành Pro) gây nên đột biến sai nghĩa.

C: thay đổi một base (C thành G) sinh ra một codon vô nghĩa (UGA).

D: thêm một base (C) gây nên đột biến acridine hoặc đột biến dịch khung, có sự dời khung đọc.

E: mất một base (A), gây nên đột biến acridine, chấm dứt đọc.

5. Lý thuyết trung tâm của sinh học phân tử

Tổng hợp protein trong tế bào có các đặc điểm sau:

- Các phân tử thông tin như nucleic acid và protein được tổng hợp theo khuôn. Tổng hợp theo khuôn vừa chính xác vừa ít tốn enzyme. Tuy nhiên, căn cứ vào hàng loạt tính chất hóa học các protein không thể làm khuôn mẫu cho sự tổng hợp của chính chúng. Vì vậy, khuôn mẫu để tổng hợp nên protein không phải là protein.

- Sinh tổng hợp protein tách rời về không gian với chỗ chứa DNA. Nhiều nghiên cứu cho thấy tổng hợp protein có thể xảy ra khi không có mặt DNA. Sự kiện này thể hiện rõ ràng nhất ở những tế bào eukaryote. Trong những tế bào này, hầu như toàn bộ DNA tập trung ở nhiễm sắc thể trong nhân, còn tổng hợp protein chủ yếu diễn ra ở tế bào chất. Tảo xanh đơn bào *Acetabularia* khi bị cắt mất phần chứa nhân vẫn tổng hợp được protein và sống vài tháng nhưng mất khả năng sinh sản. Rõ ràng, nơi chứa DNA mang thông tin di truyền và chỗ sinh tổng hợp protein tách rời nhau về không gian.

- DNA không phải là khuôn mẫu trực tiếp để tổng hợp protein, do đó phải có chất trung gian chuyển thông tin từ DNA ra tế bào chất và làm khuôn để tổng hợp protein. Chất đó phải có cả trong nhân và tế bào chất với số lượng phụ thuộc vào mức độ tổng hợp protein.

- Chất trung gian đó được xem chính là RNA nhờ các đặc điểm sau:

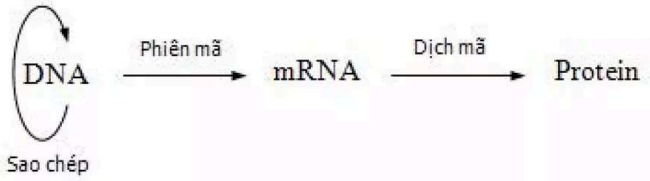
+ RNA được tổng hợp ngay ở trong nhân có chứa DNA, sau đó nó đi vào tế bào chất cho tổng hợp protein.

+ Những tế bào giàu RNA tổng hợp protein nhiều hơn.

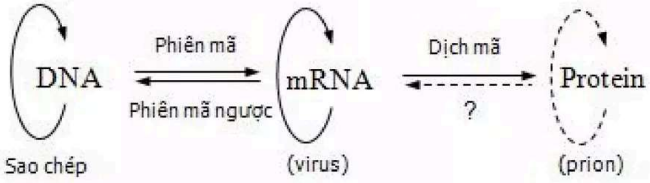
+ Về phương diện hóa học RNA gần giống DNA: chuỗi polyribo-nucleotide thẳng cũng chứa 4 loại ribonucleotide A, G, C và uracil (U). Nó có thể nhận được thông tin từ DNA qua bắt cặp bổ sung.

Nói chung, trong tế bào không thể tìm thấy chất nào khác ngoài RNA có thể đóng vai trò trung gian cho tổng hợp protein. Mối quan hệ này chính là thông tin di truyền đi từ DNA qua RNA rồi đến protein và được biểu diễn ở hình 3.3. Mối quan hệ này còn được gọi là **lý thuyết trung tâm (central dogma)**, được Crick đưa ra từ 1956 đến nay về căn bản vẫn đúng.

Vào những năm 1970, người ta đã phát hiện quá trình phiên mã ngược từ RNA tổng hợp nên DNA nhờ enzyme reverse transcriptase. Đến nay, việc sao chép (tổng hợp) RNA trên khuôn mẫu RNA cũng đã được chứng minh ở nhiều loại virus. Ngoài ra, thông tin từ protein cũng có thể được truyền sang protein (prion của bệnh bò điên). Riêng dòng thông tin từ protein ngược về mRNA/DNA thì chưa được tìm thấy (Hình 3.4).



Hình 3.3. Lý thuyết trung tâm của Crick



Hình 3.4. Những bổ sung mới vào lý thuyết trung tâm của Crick

6. DNA và mã di truyền

Chúng ta đã biết có sự liên quan đồng tuyến tính giữa DNA và phân tử protein, từ đó dễ dàng dự đoán rằng trình tự đặc hiệu của các amino acid trên phân tử protein sẽ được mã hóa bằng nhóm các nucleotide trên phân tử DNA. Có tất cả 4 loại base, nếu các base có nhóm đôi tức 2 nucleotide mã hóa cho một loại amino acid thì tất cả chỉ có 16 tổ hợp, không đủ cho 20 loại amino acid. Như vậy, đơn vị mã hóa (codon) phải gồm 3 nucleotide (xem phần 4.2).

Vấn đề tiếp theo là xác định chính xác các codon nào mã hóa cho từng amino acid. Nirenberg và Matthaei đã sử dụng enzyme để tổng hợp nhân tạo RNA. Khi dùng chỉ một loại nucleotide là U sẽ nhận được RNA là poly(U), nếu chỉ dùng A sẽ nhận được poly(A).

Năm 1961, Nirenberg và Matthaei đã dùng poly(U) thay cho khuôn mẫu mRNA để tổng hợp protein trong hệ thống vô bào (có amino acid, enzyme tổng hợp protein, nhưng không có DNA...), sản phẩm thu được là chuỗi polypeptide polyphenylalanine chỉ chứa một loại amino acid là phenylalanine. Điều đó chứng tỏ codon UUU mã hóa cho phenylalanine. Đây là codon đầu tiên được xác định. Sau đó, họ cũng chứng minh được rằng AAA mã hóa cho lysine, GGG cho glycine và CCC cho proline.

Năm 1964, Khorana tìm ra phương pháp tổng hợp mRNA nhân tạo với trình tự lặp lại (như AAG AAG AAG...) và nhờ nó giải quyết xong các vấn đề còn chưa rõ ràng.

Bảng mã di truyền (Bảng 3.4) cho thấy trong 64 codon, có 3 codon UAA, UAG, UGA không mã hóa cho amino acid được gọi là vô nghĩa (non-sense), đồng thời là codon kết thúc (termination) tức dấu chấm câu, chấm dứt chuỗi polypeptide.

Mã di truyền có tính suy biến (degeneration) tức một amino acid có nhiều codon mã hóa, chỉ trừ methionine và tryptophane chỉ có một codon (tương ứng là ATG và TGG). Các codon đồng nghĩa tức mã hóa cho cùng một amino acid thường có hai base

đầu tiên giống nhau, nhưng khác nhau ở cái thứ ba. Ví dụ: CCU, CCC, CCA và CCG tất cả đều mã hóa cho proline. Trên thực tế, U và C luôn luôn tương đương nhau ở vị trí thứ ba, còn A và G tương đương nhau trong 14 trên 16 trường hợp.

Trừ một số ngoại lệ, mã di truyền có tính phổ biến (universal) tức toàn bộ thế giới sinh vật có chung bộ mã di truyền.

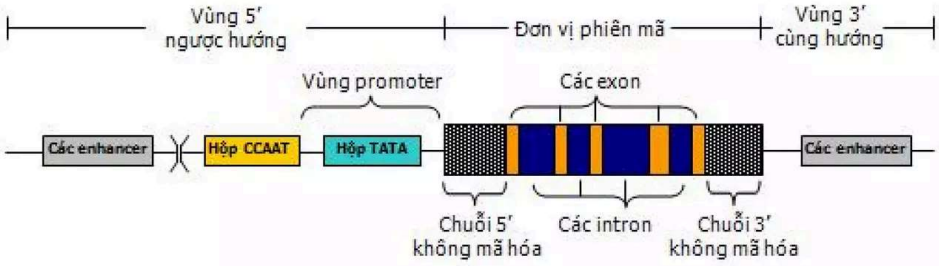
III. Cấu trúc và chức năng của gen

Khi nghiên cứu các quy luật di truyền Mendel và học thuyết di truyền nhiễm sắc thể, gen được quan niệm như một điểm trên nhiễm sắc thể, vừa là đơn vị chức năng xác định một tính trạng, vừa là đơn vị đột biến, vừa là đơn vị tái tổ hợp. Cùng với sự phát triển của di truyền học, khái niệm về gen được cụ thể hóa thêm, cấu trúc và chức năng của gen được hiểu chi tiết hơn.

1. Cấu trúc gen

Một gen thường được xem gồm có hai phần, phần mang mã di truyền được phiên mã sang phân tử mRNA và phần DNA làm nhiệm vụ điều khiển hoạt động của gen (Hình 3.5). Cả hai phần đều có cấu trúc cơ bản khá giống nhau ở mọi sinh vật prokaryote và eukaryote. Tuy nhiên, gen eukaryote còn có cấu trúc đặc thù liên quan đến các cơ chế kiểm soát khác nhau đối với hoạt động của gen.

Cấu trúc một gen điển hình nói chung đều có promoter, vị trí để RNA polymerase hoạt động khởi đầu phiên mã. Đôi khi nằm rất xa gen, có thể thấy các enhancer (vùng tăng cường) hoặc silencer (vùng ức chế) có vai trò liên quan đến quá trình phiên mã. Độ dài của một gen thay đổi tùy theo số lượng và độ dài của các intron chứa trong nó.



Hình 3.5. Cấu trúc chung của một gen

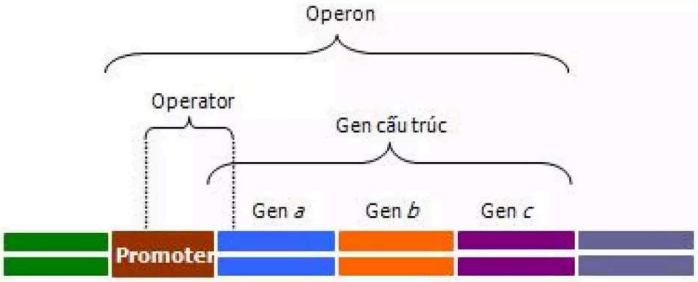
Vùng DNA mang mã di truyền sẽ được phiên mã sang phân tử mRNA. Quá trình này thực hiện theo chiều 5'→3' trên sợi mRNA đang được tổng hợp. Không phải mọi phiên mã di truyền trên phân tử mRNA đều được dịch mã sang phân tử protein. Hai đầu 5' và 3' của phân tử mRNA gồm một số nucleotide không được dịch mã mà lại liên quan đến tính bền vững của phân tử mRNA hoặc tham gia kiểm soát quá trình dịch mã. Hai đoạn này gọi là vùng không dịch mã 5' và 3' (untranslated region). Vùng không dịch mã 5' nằm trước điểm khởi đầu dịch mã (3 nucleotide AUG mã hóa cho methionine đầu tiên của chuỗi polypeptide) và vùng không dịch mã 3' nằm sau điểm kết thúc dịch mã (stop codon có thể là UAA, UGA và UAG). Do tế bào prokaryote không có cấu trúc nhân nên quá trình phiên mã (tổng hợp mRNA) và dịch mã (tổng hợp protein) xảy ra đồng thời. Còn phân tử mRNA của eukaryote được phiên mã trong nhân, sau đó phải cắt bỏ intron và gắn các exon lại, chịu biến đổi tại các đầu 5' và 3' trước khi vận chuyển ra ngoài tế bào chất để dùng làm khuôn mẫu tổng hợp protein.

Hoạt động của một gen được đánh giá thông qua quá trình phiên mã (tổng hợp mRNA) và quá trình dịch mã (tổng hợp protein). Hoạt động này được kiểm soát rất chặt chẽ bằng các cơ chế khác nhau ở mọi giai đoạn, như bắt đầu và kết thúc phiên mã, quá trình biến đổi mRNA, quyết định tính bền vững và kiểm tra lại thông tin di truyền trên các phân tử này... Do cấu trúc sắp xếp các gen prokaryote khác với gen eukaryote nên sự phối hợp giữa các cơ chế điều khiển mang tính chất riêng biệt cho từng loại genome.

Các gen prokaryote thường sắp xếp nằm gần nhau và chịu sự điều khiển chung của một promoter, tức là chúng được phiên mã sang cùng một phân tử mRNA. Cấu trúc này được gọi là operon. Như vậy, một operon gồm hai hay nhiều gen nằm cạnh nhau trên một nhiễm sắc thể. Thông thường, đó là các gen cùng tham gia vào một con đường chuyển hóa, ví dụ như các gen mã hóa cho các enzyme cần thiết cho quá trình chuyển hóa glucose.

Do có chung promoter điều khiển cho mọi gen nằm trong một operon cho nên chỉ có một loại phân tử mRNA được tổng hợp từ một operon (mang thông tin di truyền của tất cả các gen nằm trong đó). Nói cách khác, quá trình phiên mã của các gen trong một operon xảy ra đồng thời và phân tử mRNA đặc trưng cho operon được gọi là mRNA-polycistron.

Tuy nhiên, điều cần ghi nhớ là quá trình dịch mã trên các phân tử mRNA-polycistron xảy ra hoàn toàn độc lập với nhau. Mỗi đoạn tương ứng với một gen trên phân tử này đều có vị trí bám của ribosome, có mã bắt đầu và kết thúc tổng hợp chuỗi polypeptide riêng biệt. Do đó, tốc độ tổng hợp các protein trên các phân tử mRNA-polycistron hoàn toàn khác nhau (Hình 3.6).



Hình 3.6. Cấu trúc operon trong genome vi khuẩn. Một operon là một đơn vị phiên mã đơn bao gồm một chuỗi các gen cấu trúc (structural genes), một promoter và một operator.

2. Sự phân chia nhỏ của gen

Khái niệm locus được đưa ra để chỉ vị trí của gen trên nhiễm sắc thể, là vị trí của tất cả các allele của dãy đa allele. Bản thân hiện tượng đa allele cho thấy gen có cấu tạo phức tạp, sự biến đổi của gen có thể dẫn đến nhiều trạng thái allele khác nhau.

2.1. Hiện tượng allele giả

Theo quan niệm cổ điển gen là đơn vị tái tổ hợp. Nếu cá thể mang hai allele lặn a_1/a_2 của một dãy đa allele sẽ tạo thành hai loại giao tử là a_1 và a_2, lai phân tích với bố mẹ đồng hợp tử lặn sẽ chỉ cho kiểu hình đột biến a_1 và a_2 mà không có dạng tái tổ hợp hoang dại. Ví dụ:

Bố mẹ	a_1/a_2	×	a_1/a_1
Giao tử	a_1 và a_2		a_1
Các con lai	a_1/a_1 và a_2/a_1 - cả hai đều là kiểu hình đột biến		

Tuy nhiên, nhiều thí nghiệm cho thấy nếu tăng số cá thể thí nghiệm lên 10.000 hoặc 100.000, thì có thể phát hiện có dạng kiểu hình hoang dại do tái tổ hợp.

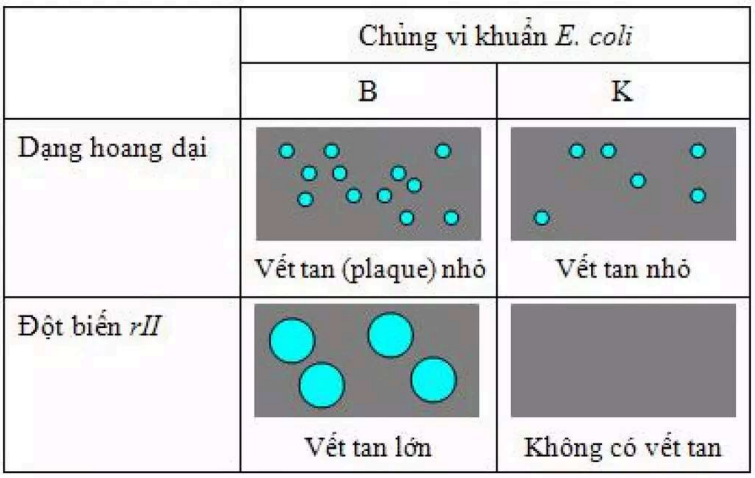
Ví dụ: Trường hợp locus mắt quả trám ở ruồi giấm, có 18 allele. Khi tăng số cá thể nghiên cứu lên nhiều lần, người ta phát hiện các allele xếp thành 3 nhóm A, B và C. Các allele của cùng một nhóm, khi lai lẫn nhau, không cho kiểu hình tái tổ hợp hoang dại mắt bình thường, mà chỉ có kiểu hình mắt quả trám. Nhưng lai allele của nhóm này với allele của nhóm khác sẽ có xuất hiện kiểu hình hoang dại do tái tổ hợp. Hiện tượng này được gọi là allele giả.

Hiện tượng allele giả cho thấy gen phân chia nhỏ về mặt tái tổ hợp, có thể xảy ra tái tổ hợp giữa các phần trong gen. Lúc đầu, hiện tượng allele giả được coi là trường hợp ngoại lệ, nhưng khi tăng số cá thể nghiên cứu lên nhiều lần thì rõ ràng đó là hiện tượng

phổ biến. Nó được tìm thấy ở nhiều đối tượng khác nhau như nấm men *S. cerevisiae*, ngô, bồ câu, chuột, bacteriophage...

2.2. Locus *rII* của bacteriophage T_4

Nghiên cứu chi tiết về các đột biến *rII* của bacteriophage T_4 đã làm sáng tỏ hơn về cấu trúc gen. Bacteriophage T_4 ở dạng hoang dại r^+ có khả năng xâm nhiễm đồng thời hai chủng *E. coli* B và K, trong khi các đột biến *rII* chỉ xâm nhiễm chủng B mà không xâm nhiễm chủng K (Hình 3.7).



Hình 3.7. Kiểu hình của các đột biến *rII* của phage T_4

Benzer (1955) đã thu được vài nghìn đột biến *rII* có nguồn gốc độc lập với nhau. Ông cho lai các đột biến này với nhau và căn cứ vào sự xuất hiện các dạng tái tổ hợp hoang dại r^+ mà lập bản đồ các điểm đột biến (mutation sites).

Trước thí nghiệm của ông, *rII* được coi là một locus. Tuy nhiên, thí nghiệm của ông đã cho thấy các đột biến xếp thành hai nhóm *rII*A và *rII*B. Lai các đột biến *rII*A × *rII*B sẽ có r^+, nhưng lai *rII*A × *rII*A và *rII*B × *rII*B thì kiểu hình đột biến là *r*.

Cho đến nay, chúng ta định nghĩa một gen là nhờ dựa trên kiểu hình đột biến và vị trí trên bản đồ của nó. Bacteriophage là một mô hình di truyền đơn giản (genome của *E. coli* dài khoảng 4.600.000 bp, trong khi bacteriophage T_4 là 165.000 bp và bacteriophage λ khoảng 46.500 bp), chúng có thể sinh sản một số lượng lớn rất nhanh (10^{10} hoặc hơn thế) và dễ dàng phân tích. Các thí nghiệm thực hiện với đột biến *rII* của T_4 được thiết lập dựa trên cơ sở sau:

- Các gen có một phạm vi và ranh giới hạn chế.

- Các gen có thể chia được, có thể có sự tái tổ hợp giữa hai allele trong một gen đơn.

- Hoạt động của gen có thể được phân tích bởi sự phân tích bổ sung.

Kết quả thí nghiệm cho thấy, gen có thể phân chia nhỏ về mặt đột biến. Các đoạn *rII*A và *rII*B được gọi là cistron, đơn vị chức năng nhỏ nhất đảm bảo khả năng xâm nhiễm chủng K. Thuật ngữ cistron thực chất là gen, ngày nay nó chỉ có tính chất lịch sử, ít được sử dụng. Theo quan niệm hiện nay, *rII*A và *rII*B là hai locus. Hai khái niệm mới được đưa ra là muton-đơn vị đột biến và recon-đơn vị tái tổ hợp.

Benzer đã tìm thấy 2.000 điểm đột biến trên đoạn gen được nghiên cứu, chúng phân bố không đều nhau, có những điểm tập trung nhiều đột biến hơn. Chiều dài gen khoảng 900 nucleotide. Đơn vị đột biến muton ở đây tương ứng với 900/2.000. Số đột biến ghi nhận có thể thấp hơn so với thực tế nên muton có thể tương ứng với một cặp nucleotide. Giống như vậy recon có thể tương ứng với một cặp nucleotide.

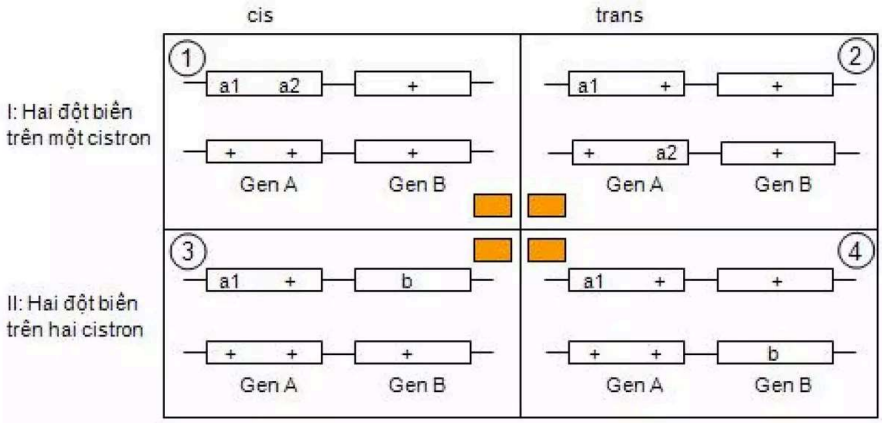
Tóm lại, gen là đơn vị chức năng, có thể chia nhỏ bởi các đơn vị đột biến tái tổ hợp.

3. Thử nghiệm chức năng allele

Muốn nghiên cứu cấu trúc bên trong một gen, phải tìm hiểu nhiều allele của gen đó. Nhiều đột biến có kiểu hình giống nhau nhưng không allele với nhau. Thử nghiệm chức năng allele được sử dụng để xác định xem hai đột biến có allele với nhau không. Đây chính là thử nghiệm mà Benzer dùng để lập bản đồ locus *rII*.

Thử nghiệm này còn được gọi là thử nghiệm bổ sung (complementary test) vì nó cho biết sai hỏng chức năng ở hai đột biến có bổ sung tức bù trừ cho nhau được không.

Phương pháp thử này cũng được gọi là thử nghiệm đều-lệch (*cis-trans* test). Sở dĩ như vậy là vì phép thử nghiệm này so sánh hiệu quả kiểu hình của các gen đột biến ở hai vị trí khác nhau trên nhiễm sắc thể tương đồng. Ở vị trí lệch (*trans*) các đột biến nằm trên hai nhiễm sắc thể, còn ở vị trí đều (*cis*) các đột biến nằm trên cùng một nhiễm sắc thể. Trường hợp sai hỏng ở hai gen khác nhau nên có thể bổ sung được, còn trường hợp sai hỏng ở cùng một gen không bù đắp được sẽ dẫn đến kiểu hình đột biến.

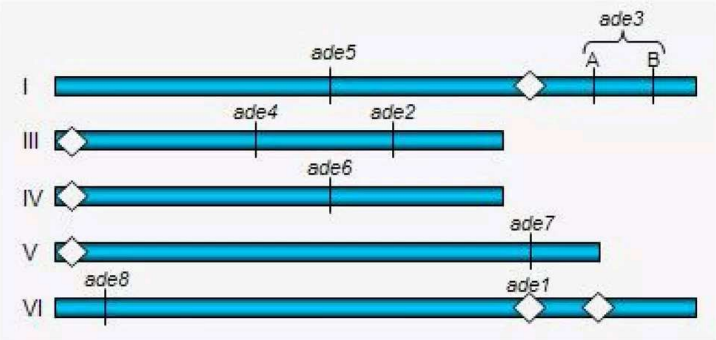


Hình 3.8. Thử nghiệm chức năng allele. I: có kiểu hình đột biến do sai hỏng cùng một gen nên không bù đắp được. II: có kiểu hình hoang dại do sai hỏng khác gen nên bù trừ được cho nhau.

4. Gen là đơn vị chức năng nhỏ nhất

Thử nghiệm chức năng allele có thể được thực hiện dễ dàng trên các đối tượng vi sinh vật với các đột biến hóa sinh, thường là các đột biến khuyết dưỡng (auxotroph mutant: mất khả năng tổng hợp một chất nào đó). Ví dụ: Ở nấm mốc *Neurospora crassa* có nhiều đột biến mất khả năng tổng hợp adenine (Ade^-). Các đột biến này dễ phát hiện vì có khuẩn lạc màu đỏ. Có hai dạng đột biến Ade_x và Ade_y, nếu dị hợp tử Ade_x/Ade_y có kiểu hình đột biến tức là cho khuẩn lạc màu đỏ, thì Ade_x và Ade_y là hai allele của một gen. Thử nghiệm chức năng allele cho thấy các đột biến *Ade* ở *N. crassa* tạo thành 9 nhóm. Điều đó chứng tỏ có 9 gen tổng hợp adenine ở loài nấm này: ade_1, ade_2, ade_3... trong đó ade_3 có hai locus nằm kề sát nhau là ade_3A và ade_3B (Hình 3.9).

Qua nghiên cứu các gen sản xuất adenine người ta nhận thấy quá trình tổng hợp adenine có liên quan đến 9 nhóm đột biến của gen này hoặc gen kia, và đều cùng có một kết quả là mất khả năng tổng hợp adenine. Như vậy, khái niệm gen không chỉ kiểm tra di truyền cả chu trình tổng hợp adenine, mà gen là đơn vị chức năng nhỏ nhất, kiểm tra một giai đoạn cơ bản nào đó của chu trình. Nếu biến đổi di truyền làm sai hỏng chức năng đó sẽ dẫn đến xuất hiện đột biến mất khả năng tổng hợp adenine.



Hình 3.9. Vị trí các gen *ade* trên các nhiễm sắc thể của nấm mốc *Neurospora crassa*

Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. *NXB Giáo dục,* Hà Nội.

2. Phạm Thành Hổ. 2003. Di truyền học. *NXB Giáo dục,* Hà Nội.

3. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

4. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

5. Pierce BA. 2003. Genetics: A Conceptual Approach. *W.H. Freeman & Co*. New York, USA.

6. Watson JD, Hopkins NH, Roberts JW and Weiner AM. 2004. Molecular Biology of the Gene. *The Benjamin/Cummings Publishing Company, Inc*. California, USA.

[1]Hai mươi amino acid được tìm thấy trong các phân tử protein là: Alanine (Ala), Arginine (Arg), Asparagine (Asn), Aspartic acid (Asp), Cysteine (Cys), Glutamic acid (Glu), Glutamine (Gln), Glycine (Gly), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Phenylalanine (Phe), Proline (Pro), Serine (Ser), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr) và Valine (Val).

Chương 4

Tái bản DNA

I. Chứng minh tái bản DNA theo cơ chế bán bảo thủ

1. Cơ chế tái bản bán bảo thủ

1.1. Cơ chế tái bản ở prokaryote

Đặc điểm cơ bản của sự tái bản đó là tái bản theo phương thức bán bảo thủ (semiconservative replication). Tái bản bán bảo thủ nghĩa là trong hai chuỗi của tất cả các phân tử DNA bao giờ cũng có:

- Một chuỗi của DNA cũ (từ một trong hai chuỗi của DNA mẹ).
- Một chuỗi của DNA mới (mới được tổng hợp).

Mỗi một lần tái bản đều có sự tách rời của hai chuỗi của DNA mẹ, đồng thời mỗi chuỗi mẹ tiến hành sao chép để cho một chuỗi con, chuỗi này sau đó lại kết hợp với chuỗi mẹ.

Vị trí mở xoắn kép và tổng hợp DNA mới cùng một lúc trên DNA gọi là chạc ba tái bản (replication fork) do cấu trúc của vùng tái bản có hình chữ Y. Sự tổng hợp DNA mới gắn liền với việc mở xoắn DNA cũ.

1.2. Cơ chế tái bản ở eukaryote

Sự tái bản ở tế bào eukaryote phức tạp hơn so với ở tế bào prokaryote nhưng cơ chế của sự tái bản ở eukaryote cũng tương tự như ở prokaryote và tiến hành theo các nguyên tắc:

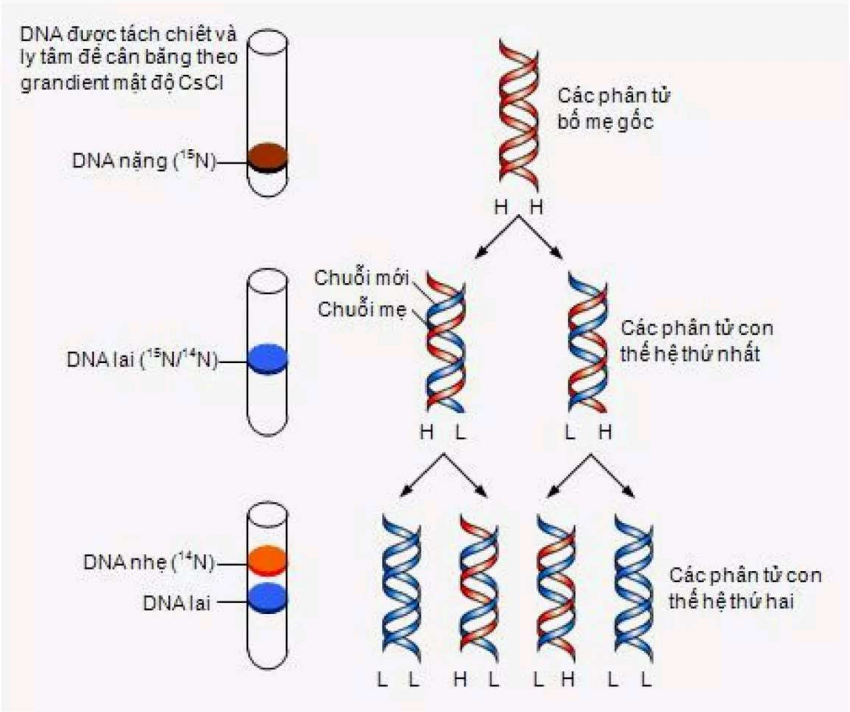
- Hai hướng.
- Bổ sung, đối song song, theo chiều 5’→ 3’.
- Không liên tục ở một trong hai chuỗi.
- Cần những RNA primer.

Tuy nhiên, có một số điểm khác như sau:

- Trong khi ở prokaryote chỉ có một điểm khởi đầu, thì sự tái bản ở eukaryote bắt đầu cùng một lúc ở nhiều điểm khởi đầu. Điều này là cần thiết do DNA của eukaryote có chiều dài rất lớn.
- Vận tốc phát triển của chạc ba tái bản ở eukaryote (khoảng 50 nucleotide/s) chỉ bằng 1/10 so với ở *E. coli*.

2. Thí nghiệm của Meselson và Stahl

Những thí nghiệm của Meselson và Stahl (1957) đã chứng minh lý thuyết tái bản DNA theo kiểu bán báo thủ (Hình 4.1). Các tác giả trên đã nuôi cấy *E. coli* trong nhiều thế hệ trong một môi trường chứa $^{15}NH_4Cl$ làm nguồn cung cấp nitrogen duy nhất. Bằng cách này, DNA được tổng hợp có ^{15}N (^{15}N là một chất phóng xạ nặng hơn chất phóng xạ thông thường ^{14}N). Ở một thời điểm nhất định (thời điểm 0), các tác giả này đã chuyển nuôi cấy vào một môi trường chứa $^{14}NH_4Cl$. Tiếp đến, sau từng thời gian đều đặn, họ phân tích DNA chiết xuất từ vi khuẩn bằng phương pháp ly tâm theo gradient CsCl.



Hình 4.1. Minh họa sự tái bản bán bảo thủ. Sơ đồ trình bày sự hợp thành các sợi đôi DNA sau 0, 1, và 2 vòng sao chép. H: chuỗi nặng (^{15}N), L: chuỗi nhẹ (^{14}N).

Kết quả thực nghiệm cho thấy:

- Ở thời điểm 0: chỉ có một phân tử tương ứng với DNA nặng ^{15}N.

- Sau một thế hệ trong môi trường chứa ^{14}N: những phân tử DNA gồm một chuỗi nặng ^{15}N (chuỗi mẹ) và một chuỗi nhẹ ^{14}N (mới được tổng hợp).

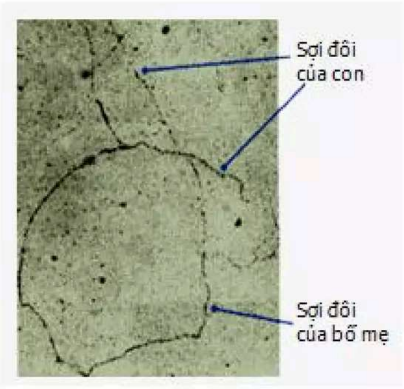
- Sau hai thế hệ trong môi trường chứa ^{14}N: có hai phân tử lai (gồm một chuỗi nặng và một chuỗi nhẹ) và hai phân tử đều gồm những chuỗi nhẹ không có chuỗi nặng.

II. Mô hình tái bản DNA-chạc ba tái bản

1. Mô hình tái bản

Mô hình tái bản được nghiên cứu trên thể nhiễm sắc của vi khuẩn *E. coli,* DNA có dạng mạch vòng sợi đôi (Hình 4.2). Để tự tái bản DNA phải tháo ra đơn giản ở một vị trí nhất định và nơi đó xuất hiện chạc ba tái bản (Hình 4.3).

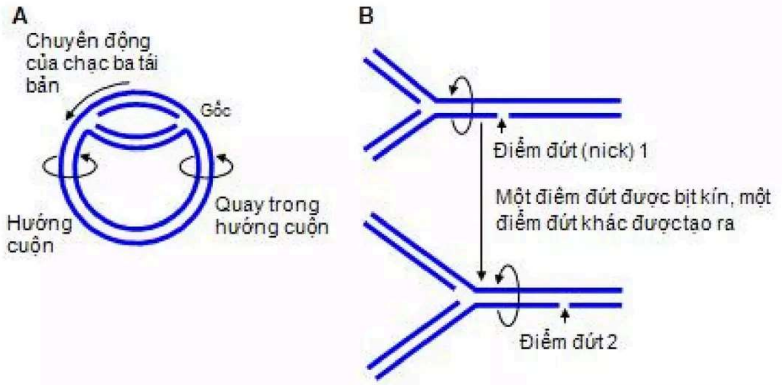
Thí nghiệm của Cairns. Sử dụng nucleotide được đánh dấu bằng đồng vị phóng xạ trong môi trường đang phân chia, ta sẽ biết được tiến trình tái bản do hạt bạc xuất hiện dưới kính hiển vi điện tử.



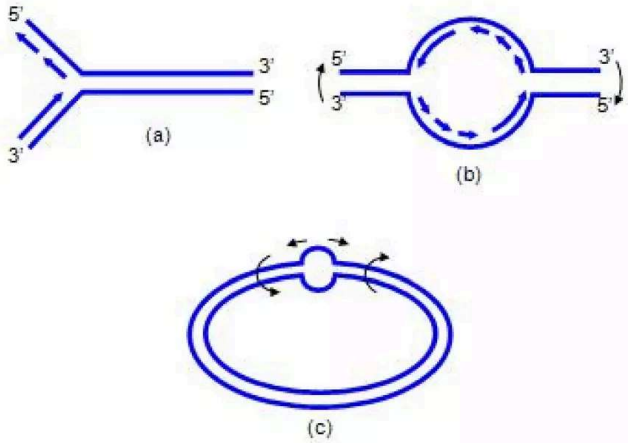
Hình 4.2. DNA dạng mạch vòng sợi đôi. Chiều dài thực tế 1,6 mm ($4{,}7\times10^6$ bp).

2. Chạc ba tái bản

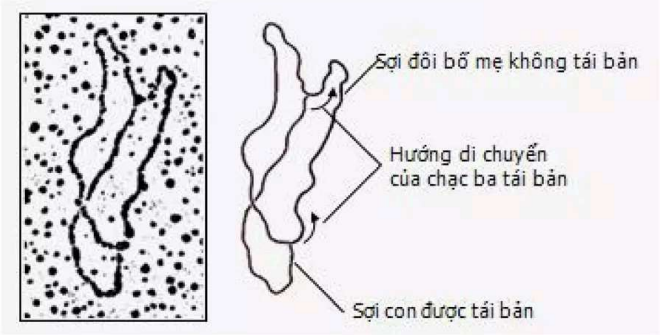
Hình 4.4 mô tả cấu trúc của chạc ba tái bản. Nhờ vào phương pháp phóng xạ ảnh tự ghi, người ta nhận thấy sự tái bản thực hiện theo hai hướng (bidirectional synthesis). Đồng thời cũng chứng minh được vi khuẩn *E. coli* (prokaryote) có duy nhất một điểm gốc tái bản, đó là điểm mà hai chuỗi xoắn kép của DNA mẹ được tách ra, tương ứng với hai chạc ba tái bản phát triển ngược chiều nhau. Chuỗi DNA sau khi tách ra được dùng làm khuôn mẫu cho sự tổng hợp DNA mới (Hình 4.5).



Hình 4.3. Sự tái bản của phân tử DNA sợi đôi mạch vòng của *E. coli*. A: sự chuyển động không cuộn lại của các nhánh trong quỹ đạo tái bản, không có các vị trí quay tự do, gây ra sự cuộn lại quá chặt của phần không được tái bản. B: cơ chế sợi đơn bị đứt (nick) phía trước của chạc ba tái bản cho phép sự quay xảy ra.



Hình 4.4. Sơ đồ cấu trúc của các chạc ba tái bản. (a): Chạc ba đơn trình bày sợi chủ (leading strand) được tổng hợp liên tục và sợi thứ (lagging strand) được tổng hợp gián đoạn. (b): Chạc ba đôi, phổ biến trong hầu hết mọi sự tái bản DNA của genome. (c): Các hướng hình học của sự tái bản DNA, mũi tên ngắn chỉ sự chuyển động dịch mã của chạc ba, mũi tên dài và cong chỉ sự quay vòng DNA cần thiết quanh các chạc ba.

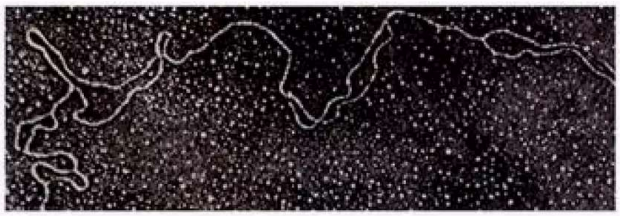


Hình 4.5. Hình ảnh dưới kính hiển vi điện tử. Phân tử DNA mạch vòng nhỏ của *E. coli* có chiều dài thực tế 0,01 mm (3.000 bp) tái bản bằng kiểu θ. Các đoạn DNA bố mẹ và con được trình bày trong hình vẽ.

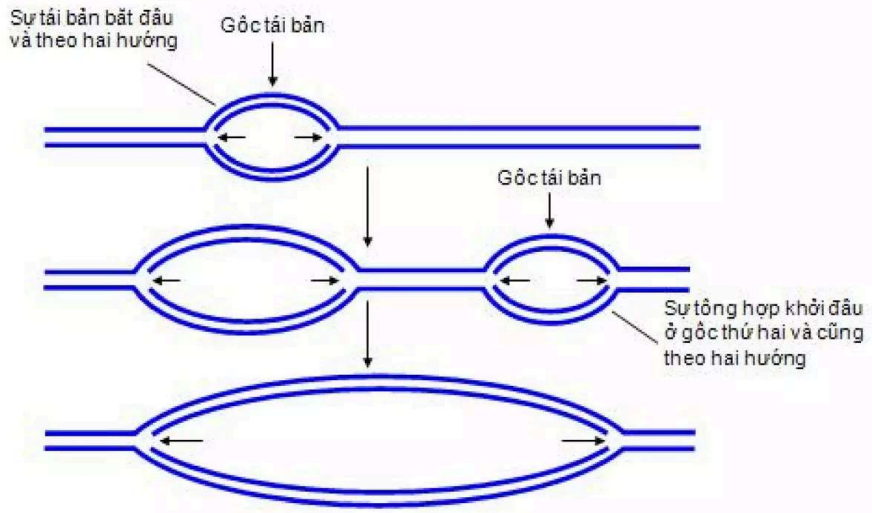
Hình 4.6 so sánh sự khác nhau giữa tái bản DNA không định hướng và tái bản theo hai hướng. Trong tái bản không định hướng, chỉ có một chạc ba tái bản. Trong khi tái bản theo hai hướng yêu cầu hai chạc ba tái bản. Mũi tên cong chỉ hướng chuyển động của các chạc ba. Hầu hết DNA tái bản theo hai hướng.

Hình 4.6. Tái bản DNA không định hướng và theo hai hướng

Hình 4.7 và 4.8 mô tả phương thức hợp nhất các vòng tái bản DNA ở ruồi giấm (*D. melanogaster*). Quá trình tái bản diễn ra đồng thời trên hàng chục ngàn vị trí khác nhau của phân tử DNA và tạo thành các vòng tái bản, các vòng tái bản sau đó sẽ mở rộng theo hai hướng để cuối cùng hợp nhất với nhau tạo thành hai phân tử DNA.



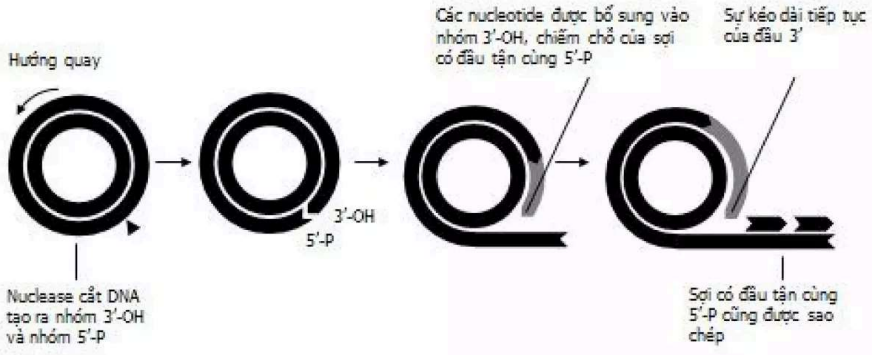
Hình 4.7. Hình ảnh dưới kính hiển vi điện tử của một đoạn nucleotide ở ruồi giấm. Phân tử DNA sợi đôi dài 30 kb cho thấy có 7 vòng tái bản.



Hình 4.8. Phương thức hợp nhất các vòng tái bản DNA của ruồi giấm. Hai gốc tái bản được trình bày trên hình vẽ, các mũi tên nhỏ chỉ hướng chuyển động của các chạc ba tái bản.

3. Tái bản DNA theo vòng tròn quay

Trường hợp bacteriophage λ (thực khuẩn thể λ) có vật chất di truyền là một phân tử DNA mạch thẳng sợi đôi, khi ta chuyển chúng vào vi khuẩn thì các đầu dính kết DNA (*cos*) của nó gắn lại theo dạng vòng tròn. Sự dính kết lại này là do hoạt động của enzyme DNA ligase giúp tạo lại dạng xoắn. Khi đó, sự tái bản DNA tiến hành theo cơ chế vòng tròn quay (Hình 4.9).



Hình 4.9. Tái bản vòng tròn quay ở bacteriophae }. DNA được tổng hợp mới có màu nhạt. Sợi thay thế được tái bản trong các đoạn ngắn.

III. Bản chất xoắn của DNA-Các giai đoạn của sự tái bản

Hình 4.10 mô tả toàn bộ quá trình tái bản DNA. Quá trình này trải qua ba giai đoạn chính sau:

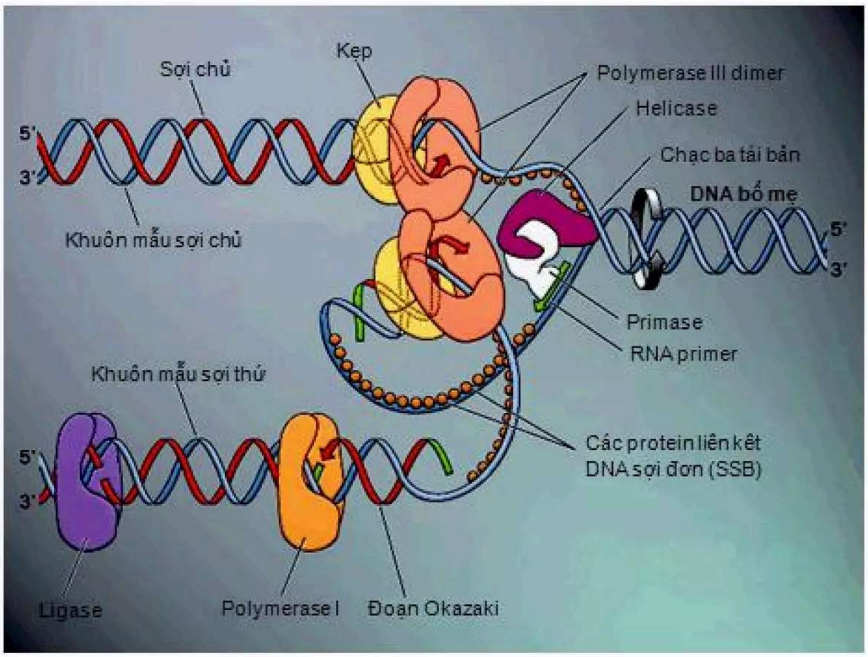
1. Mở xoắn

Trước tiên ta thấy quá trình mở xoắn của hai sợi DNA cần thiết phải có một enzyme rất quan trọng đó là helicase (còn gọi là enzyme mở xoắn).

Chẳng hạn, trên phân tử *E. coli* chỉ có một gốc tái bản (ký hiệu là *ori* C) dài 245 bp. Có ít nhất 8 enzyme hoặc protein tham gia vào giai đoạn khởi đầu của sự tái bản. Những enzyme-protein này mở xoắn DNA ở gốc *ori* C và thiết lập một phức hợp tiền mồi (prepriming complex) để chuẩn bị cho những phản ứng của giai đoạn sau.

Một phức hợp khoảng 20 protein Dna A được kết hợp với một vùng của *ori* C bắt đầu cho quá trình mở xoắn và khởi đầu sự tái bản. Phản ứng này cần ATP và một protein giống như histone của vi khuẩn (HU). Tiếp đó, protein Dna C giúp protein Dna B gắn vào vùng *ori* C, Dna B chính là helicase sẽ mở xoắn DNA theo hai hướng để tạo ra hai

chạc ba tái bản. Các phân tử protein liên kết sợi đơn (single stranded binding protein, protein SSB) gắn vào chuỗi đơn DNA để ổn định chuỗi này. Enzyme gyrase (DNA topoisomerase II) mở xoắn để tạo siêu xoắn trái (−). Enzyme primase (protein Dna G) sẽ xúc tác tổng hợp RNA primer để gắn vào khuôn mẫu DNA.



2. Kéo dài-Tổng hợp chuỗi Okazaki

Giai đoạn kéo dài bao gồm sự tổng hợp cùng một lúc hai chuỗi DNA. Một chuỗi được tái bản cùng hướng phát triển của chạc ba sẽ liên tục (sợi chủ), còn chuỗi kia sẽ không liên tục bao gồm các đoạn Okazaki (sợi thứ). Các nucleotide gắn vào đầu 3' tự do, và kéo dài chuỗi ra nhờ enzyme DNA polymerase III theo chiều 5'→3'. Trong giai đoạn này, nhiều enzyme đã tham gia vào sự tổng hợp hai chuỗi DNA tại chạc tái bản. DNA helicase tiếp tục tách hai chuỗi DNA mẹ, DNA gyrase mở xoắn, protein SSB ổn định những chuỗi DNA đơn đã được tách ra, DNA ligase gắn các đoạn Okazaki trên sợi thứ.

Sinh tổng hợp DNA trong tế bào cùng một lúc diễn ra trên rất nhiều vị trí (20.000-60.000) suốt chiều dài khổng lồ của DNA khuôn mẫu. Ở các vị trí đó, enzyme helicase giúp DNA chuyển từ dạng siêu xoắn sang dạng dãn (hai sợi DNA tách ra) với tốc độ 10 μm/giờ. Với tốc độ này, để helicase có thể chạy suốt chiều dài chuỗi DNA trong một tế bào cây đậu ngựa (DNA dài tổng cộng khoảng 9 m) phải mất hơn một thế kỷ. Trong thực tế, công việc này chỉ mất khoảng 15-40 giờ.

Chuỗi DNA xoắn kép như vậy được tách đôi ở các vị trí sẽ xảy ra sinh tổng hợp, hình thành các chạc ba. Ở chạc ba, cả hai chuỗi DNA mạch đơn đều được sử dụng làm

khuôn mẫu cùng một lúc. Trên một chuỗi, sinh tổng hợp sẽ diễn ra theo chiều từ 5'→3' cùng chiều với chiều phát triển của chạc tái bản. Trên sợi còn lại, sinh tổng hợp vẫn diễn ra theo chiều 5'→3' nhưng ngược chiều với chiều phát triển của chạc tái bản, và chỉ thực hiện được từng đoạn ngắn độ 1.000 nucleotide (đoạn Okazaki). Trong khi di chuyển từng quãng, enzyme primase sẽ tổng hợp những đoạn RNA primer ngắn (11±1 nucleotide) để từ đó DNA được tiếp nối nhờ DNA polymerase III. Khi đoạn Okazaki mới được hoàn chỉnh, RNA primer được tách ra nhờ DNA polymerase I (hoạt tính exonulease 5'→3') và được thay thế bởi DNA cũng nhờ tác dụng của cùng enzyme. Các khe hở còn lại giữa các đoạn Okazaki được DNA ligase gắn và DNA trở lại dạng chuỗi xoắn kép: một phân tử DNA mới được hình thành (Hình 4.11).

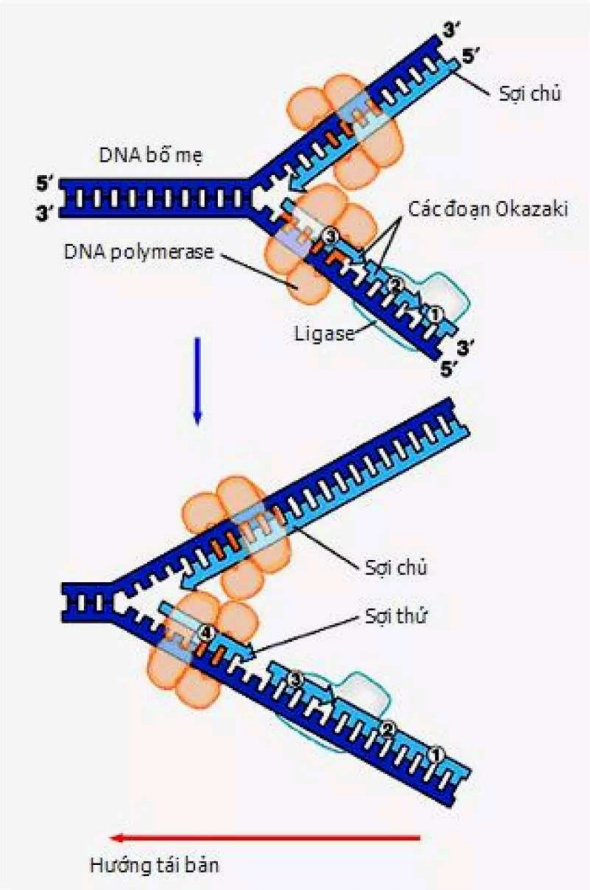
3. Kết thúc

Cuối cùng, hai chạc ba tái bản gặp nhau ở phía đối diện của nhiễm sắc thể vòng của *E. coli*. Người ta biết rất ít về những phản ứng của giai đoạn này. Có thể hoạt động của một loại DNA topoisomerase là cần thiết để tách hai phân tử DNA vòng đã được tổng hợp. Quá trình phân đôi hai phân tử DNA trong tế bào con khi phân chia tế bào cũng chưa được biết rõ.

IV. Khái niệm mồi

Tái bản DNA chỉ xảy ra khi có sự khởi động mồi. Thường mồi (primer) là đoạn RNA ngắn gắn trước đầu DNA.

Quan sát chạc ba tái bản ta thấy DNA polymerase gắn nhóm phosphate vào đầu 3'-OH tự do. Tuy nhiên, lúc đầu chưa có đầu 3'-OH tự do, do đó ở ngay điểm gốc tái bản enzyme RNA polymerase (enzyme primase) sẽ hoạt động tạo nên một đoạn mồi ngắn để có đầu 3'-OH tự do, nhờ đó DNA polymerase mới bắt đầu hoạt động tái bản.

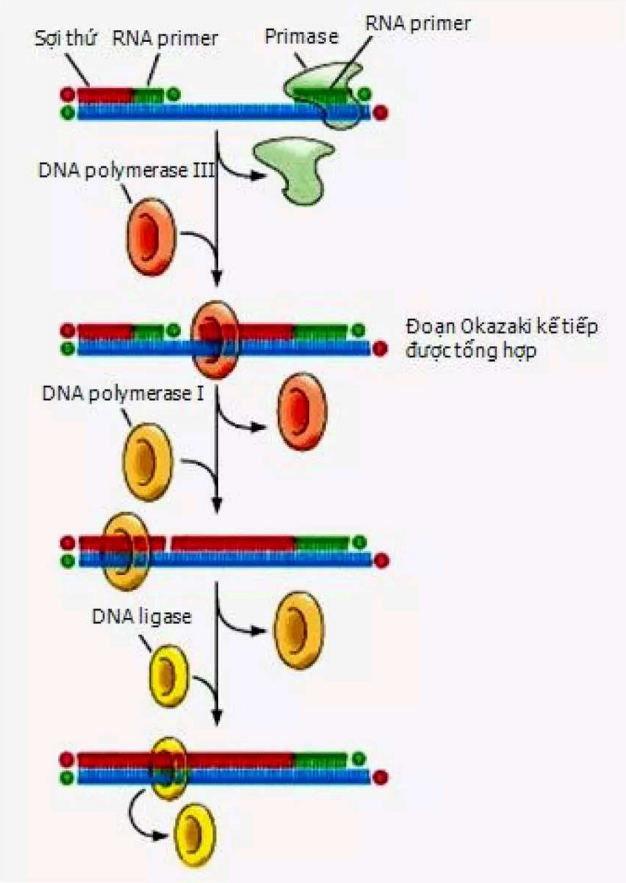


Hình 4.11. Hai sợi DNA mới được tổng hợp theo hai kiểu khác nhau

Nghiên cứu tái bản ở bacteriophage M13 (DNA một sợi vòng). Gọi DNA + (M13) là sợi (+). Từ sợi (+) tạo ra dạng tái bản (–), sợi (–) này sẽ dùng làm khuôn tạo ra sợi (+) khác.

DNA polymerase sẽ không hoạt động, nếu không có mồi ghép đôi vào sợi (+) đầu tiên. Enzyme RNA polymerase sẽ tổng hợp đoạn mồi bằng cách bổ sung một số nucleotide vào vị trí khởi đầu tái bản, tạo nên đầu 3'-OH tự do. RNA polymerase bao gồm các đơn vị α_2, β, β' và σ. Chất kháng sinh rifampicin có tác dụng ức chế đơn vị β, khi đó RNA polymerase sẽ bị ức chế và M13 không có hiện tượng tái bản. Như vậy, chính RNA polymerase giữ vai trò khởi động để sau đó DNA polymerase mới bắt đầu hoạt động tái bản được. Vì vậy, đầu tiên phải có một đoạn mồi gắn với DNA bằng liên kết cộng hóa trị, và đoạn mồi này do RNA polymerase tổng hợp nên.

Tiếp tục, đoạn mồi RNA sẽ bị loại ra bởi DNA polymerase (khác với DNA polymerase tái bản ở trên). DNA polymerase này bắt đầu gắn các nucleotide vào đầu 3'-OH tự do, và sau cùng enzyme DNA ligase nối các đoạn tái bản này lại (Hình 4.12).



Hình 4.12. Tổng hợp các đoạn Okazaki. Quá trình này đòi hỏi sự gắn mồi, kéo dài, loại bỏ RNA, làm đầy các khoảng trống và hàn các điểm đứt. Primase tổng hợp đoạn mồi RNA, DNA polymerase III kéo dài đoạn mồi RNA thành đoạn Okazaki ở sợi thứ, DNA polymerase I sử dụng dịch mã điểm đứt để thay thế đoạn mồi RNA bằng DNA, và cuối cùng DNA ligase hàn điểm đứt.

Như vậy, việc tái bản DNA cần các enzyme sau:

- Helicase: mở xoắn để tách sợi DNA đang xoắn.
- SSB: gắn trên DNA một sợi để sợi luôn luôn ở tình trạng mở.
- RNA polymerase (primase): tác động hình thành đoạn mồi RNA.
- DNA polymerase I: loại bỏ đoạn mồi RNA.
- DNA polymerase III (khác với loại trên) tác động tổng hợp DNA bằng cách kéo dài đoạn mồi RNA.
- DNA ligase: gắn các đoạn Okazaki đã tổng hợp với nhau.

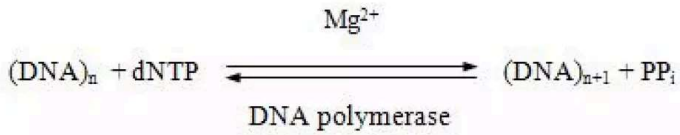
Từ các nghiên cứu về tái bản DNA của bacteriophage M13, người ta đã đưa ra sơ đồ giả thuyết về sự biến đổi DNA một sợi của M13 thành dạng sao chép tái bản (RF).

Và quan sát trên chạc ba tái bản đang phát triển, ta có thể theo dõi được sự tái bản liên tục và không liên tục của các đoạn DNA mới bổ sung cho sợi DNA khuôn mẫu.

V. Enzyme tái bản

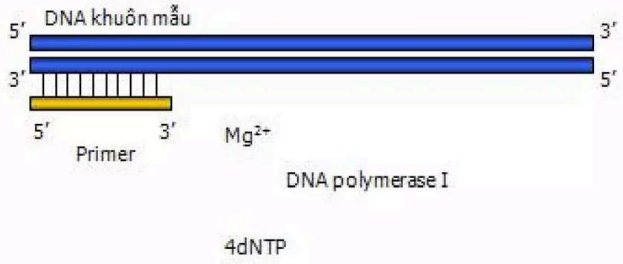
1. DNA polymerase

Đây là enzyme chủ yếu của sự tái bản, chịu trách nhiệm tổng hợp hai chuỗi DNA ở chạc tái bản. Có ba loại DNA polymerase khác nhau, được ký hiệu là I, II và II. Cơ chế và vai trò của các enzyme DNA polymerase cũng rất khác nhau.



1.1. DNA polymerase I

Do Kornberg phát hiện năm 1955. DNA polymerase I của *E. coli* mang chuỗi polypeptide có khối lượng phân tử là 109.000 Da với ba hoạt tính riêng biệt. Nhiệm vụ chính của DNA polymerase I là xúc tác cho quá trình lắp ráp các dNTP từng cái một vào đầu 3' tự do của đoạn mồi đang gắn trên DNA khuôn mẫu. Kết quả là sợi DNA mới sẽ dài dần về phía đầu 3' (Hình 4.13).



Hình 4.13. Các thành phần cần thiết cho sinh tổng hợp DNA

DNA polymerase I của *E. coli* còn có hoạt tính exonuclease (hoạt tính cắt các chuỗi nucleotide ở các đầu tự do của DNA) 3'→5' xúc tác cho sự thoái hóa bậc thang từ đầu 3'

của cả DNA sợi đôi và sợi đơn khi không có dNTPs. Trong trường hợp có dNTPs, hoạt tính exonuclease trên sợi đôi sẽ bị ức chế bởi hoạt tính polymerase. Trong quá trình tổng hợp DNA, hoạt tính exonuclease thực hiện chức năng đọc sửa bằng cách cắt bỏ những nucleotide lắp ráp sai.

Ngoài ra, enzyme này còn có hoạt tính exonulease 5'→3' xúc tác cho sự chuyển chỗ của các nucleotide từ đầu 5' của điểm đứt và bổ sung tức thời các nucleotide từ đầu 3' tạo ra chuyển động của điểm đứt dọc theo DNA.

1.2. DNA polymerase II và DNA polymerase III

Khởi đầu, enzyme DNA polymerase I được xem như giữ vai trò chủ yếu trực tiếp trong tái bản DNA. Nhưng sau đó, người ta quan sát thấy ở *E. coli* đột biến không có DNA polymerase I mà vẫn có sự tái bản, và phát hiện rằng chính DNA polymerase III giữ vai trò tái bản DNA. Nghiên cứu tính chất và so sánh ba loại enzyme trên người ta nhận thấy DNA polymerase III mới là enzyme thực sự điều khiển tổng hợp DNA. Trong khi đó DNA polymerase I chỉ có vai trò loại bỏ RNA mồi nhờ cắt đầu 5'→3' và thay vào chỗ RNA mồi bằng DNA khác, còn DNA polymerase II chưa rõ tác dụng, có lẽ nó tham gia vào quá trình sửa chữa (thay một đoạn DNA hỏng bằng một đoạn DNA bình thường) hoặc thay thế DNA polymerase I khi có khó khăn trong tổng hợp DNA.

Về tốc độ làm việc, các DNA polymerase rất khác nhau. Trong một giây, DNA polymerase I gắn được 10 dNTP trong khi DNA polymerase II chỉ gắn được 0,5 dNTP và DNA polymerase III gắn tới 150 dNTP.

2. Các topoisomer và DNA topoisomerase

2.1. Topoisomer

Giả sử có hai phân tử DNA mạch vòng có cùng trình tự nucleotide. Nhưng hai phân tử này có thể có số vòng (linking number-Lk) khác nhau trong phân tử. Số vòng ở đây được định nghĩa là số lần của một chuỗi DNA quấn xung quanh một chuỗi khác. Những trường hợp này được gọi là topoisomer. Topoisomer có các dạng sau:

2.1.1. Dạng lỏng lẻo (relaxed DNA)

Ở dạng này sức căng của xoắn kép là tối thiểu. Đó là dạng cấu trúc ổn định nhất của phân tử.

2.1.2. Dạng siêu xoắn (supercoiled DNA)

Trục của xoắn kép có thể cuộn xung quanh mình tạo thành một siêu xoắn. Có hai dạng siêu xoắn sau:

- Siêu xoắn dương (+). Số vòng tăng, xoắn kép xoắn cùng chiều (xoắn phải) tạo thành dạng siêu xoắn dương (positive supercoil).

- Siêu xoắn âm (>). Số vòng giảm, xoắn kép xoắn theo chiều ngược lại (chiều trái) tạo thành dạng siêu xoắn âm (negative supercoil),

Phần lớn những phân tử DNA trong tự nhiên ở dạng siêu xoắn âm.

2.2. DNA topoisomerase

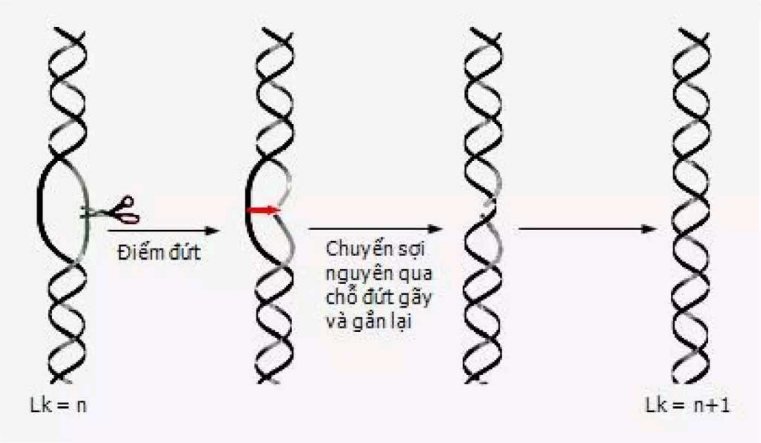
Enzyme DNA topoisomerase (là một enzyme nuclease thuận nghịch) có tác dụng thay đổi số Lk của DNA. Các DNA topoisomerase có tác dụng thêm vào hoặc loại bớt những siêu xoắn trong phân tử DNA xoắn kép. Có hai loại DNA topoisomerase sau:

2.2.1. DNA topoisomerase I

DNA topoisomerase I tồn tại ở cả prokaryote lẫn eukaryote và có tác dụng sau:

- Cắt một chuỗi DNA và tháo một vòng xoắn sau vị trí cắt.

- Nối lại chuỗi DNA đã bị cắt bởi liên kết phosphodiester mới. DNA sau khi được tái tạo lại có cấu trúc lỏng lẻo hơn (Hình 4.14).



Hình 4.14. Cơ chế hoạt động của topoisomerase I. Enzyme cắt một sợi đơn của DNA sợi đôi, chuyển sợi nguyên vẹn qua chỗ đứt gãy, sau đó hàn chỗ đứt gãy lại. Quá trình này tăng số Lk lên 1.

2.2.2. DNA topoisomerase II

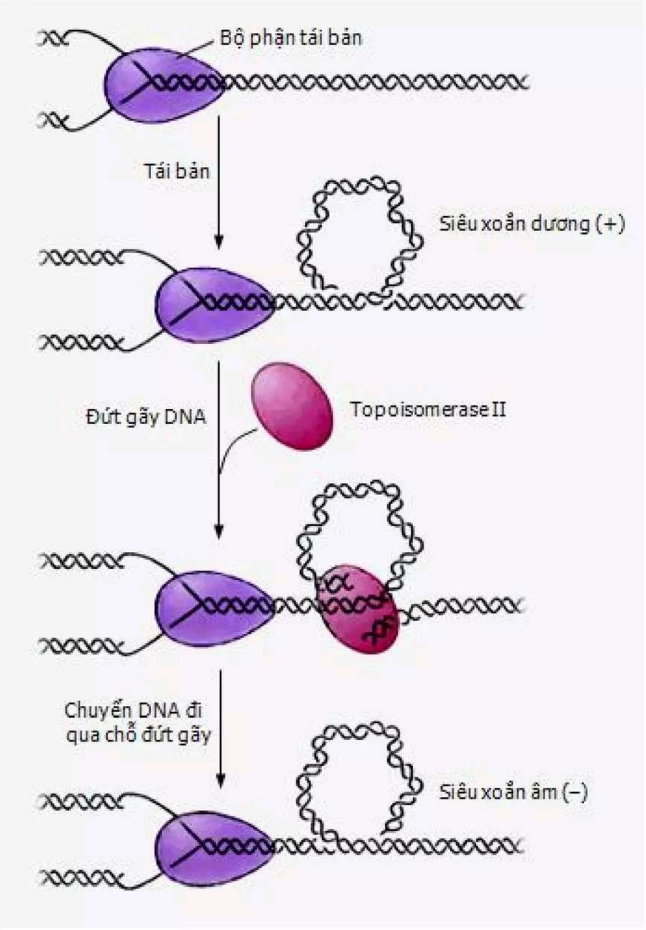
Ở prokaryote, DNA topoisomerase II có tên là DNA gyrase. DNA topoisomerase II có tác dụng cắt tạm thời hai chuỗi của DNA, có tác dụng sắp xếp lại siêu xoắn, tạo ra siêu xoắn trái (–) của chuỗi DNA xoắn kép (Hình 4.15). Phản ứng này cần 1 ATP.

Ở eukaryote, DNA topoisomerase II cũng được tìm thấy nhưng ít được nghiên cứu.

3. Helicase và protein SSB

3.1. Helicase

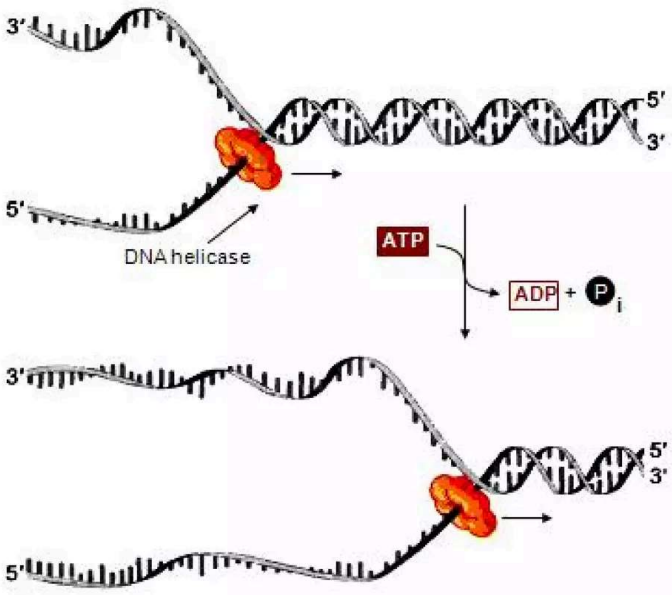
Enzyme helicase (còn có tên là enzyme deroulase) có nhiệm vụ giúp chuỗi DNA từ dạng siêu xoắn sang dạng dãn thành hai sợi đơn bằng cách cắt các liên kết hydrogen giữa những base bổ sung. Dạng cấu trúc này cần thiết cho các đoạn trên chuỗi DNA mà ở đó có nhu cầu sinh tổng hợp. Phản ứng này cần sự có mặt của ATP (Hình 4.16).



Hình 4.15. Hoạt động của topoisomerase II ở chạc ba sao chép

3.2. Protein SSB

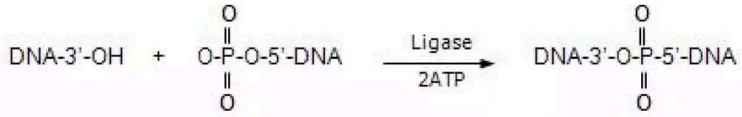
Các protein SSB liên kết phối hợp với DNA sợi đơn nhưng không liên kết với DNA sợi đôi. Chúng được xem như là các chất phản ứng trong tạo dòng phân tử xuất phát từ chỗ chúng có khả năng phá vỡ sự ổn định của các cấu trúc thứ cấp bên trong sợi nucleotide; tăng nhanh sự tái ủ của các polynucleotide bổ trợ và tăng hoạt tính của các DNA polymerase bằng cách loại bỏ các cấu trúc thứ cấp bên trong sợi là những yếu tố ngăn cản sự tiến triển của các enzyme này (Hình 4.17). Nhờ tính chất này, các protein SSB trở thành các chất phản ứng hữu ích trong xác định trình tự DNA.

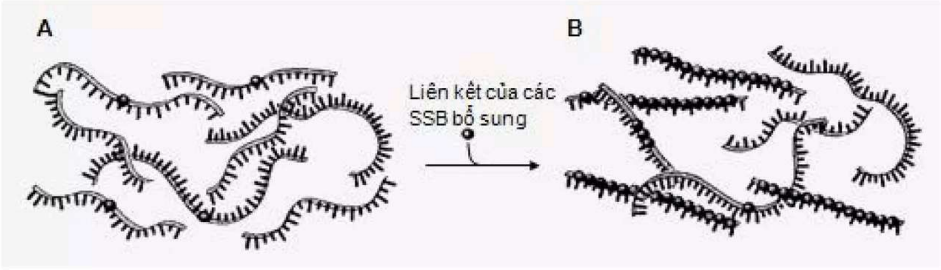


Hình 4.16. Hoạt động của enzyme helicase. DNA helicase tách rời hai sợi của chuỗi xoắn kép. Khi ATP được bổ sung vào DNA helicase được liên kết với sợi đơn, thì helicase di chuyển với một độ phân cực xác định trên DNA sợi đơn. Độ phân cực nghĩa là DNA helicase được liên kết với khuôn mẫu sợi thứ trên chạc ba tái bản.

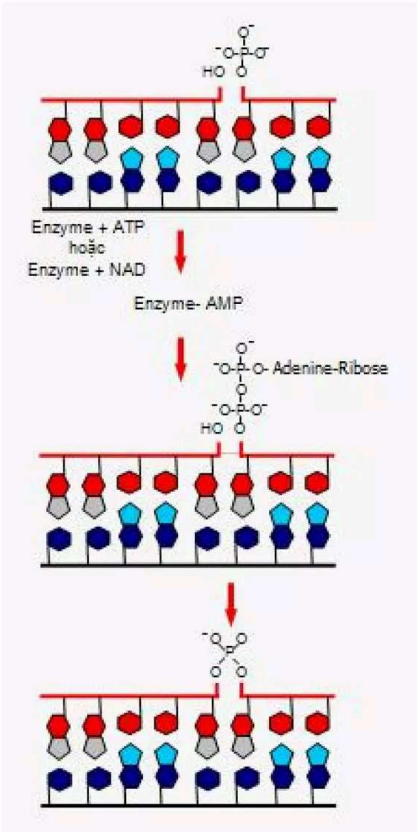
4. DNA ligase

Có nhiệm vụ nối hai đầu 3'-OH và 5'-PO_4 của hai đoạn DNA rời thành một đoạn liên tục. Một mối hàn như vậy cần một năng lượng là 2ATP (đối với eukaryote) hoặc NAD^+ (đối với vi khuẩn).





Hình 4.17. Liên kết của protein SSB lên DNA sợi đơn. Một lượng giới hạn của SSB liên kết với 4 trong 9 phân tử DNA sợi đơn. Khi bổ sung thêm protein SSB, nó sẽ liên kết với các protein SSB đã liên kết từ trước. Chỉ sau khi protein SSB bao bọc hoàn toàn các phân tử DNA sợi đơn ban đầu, thì nó mới tiếp tục liên kết với các phân tử DNA sợi đơn khác.



Hình 4.18. DNA ligase hàn các điểm đứt giữa các nucleotide gần nhau

Đặc điểm của DNA ligase là không làm việc với các sợi DNA đơn (single strand) mà chỉ hàn nối các đoạn DNA ở dạng chuỗi xoắn kép. Khi nối DNA, có thể gặp 2 trường hợp: đầu so le-đầu dính (protruding ends-cohesive ends) hoặc đầu bằng (blunt ends). Trường hợp đầu so le, các nucleotide nằm trên phần so le của hai đoạn DNA phải tương đồng thì DNA polymerase mới hoạt động được.

Tài liệu tham khảo/đọc thêm

1. Phạm Thành Hổ. 2003. Di truyền học. *NXB Giáo dục,* Hà Nội.

2. Lê Đức Trình. 2001. Sinh học phân tử của tế bào. *NXB Khoa học và Kỹ thuật*, Hà Nội.

3. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

4. Karp G. 2002. Cell and Molecular Biology: Concepts and Experiments. 3rd ed. *John Wiley & Sons, Inc*. New York, USA.

5. Lewin B. 2000. Gene VII. *Oxford University Press,* Oxford, UK.

6. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *Freeman and Company*, New York, USA.

7. Watson JD, Hopkins NH, Roberts JW and Weiner AM. 2004. Molecular Biology of the Gene. *The Benjamin/Cummings Publishing Company, Inc*. California, USA.

8. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company*, New York, USA.

Chương 5

Phiên mã

Phiên mã là quá trình tổng hợp RNA từ khuôn mẫu DNA. Quá trình này về phương diện hóa học và enzyme rất giống với quá trình tái bản DNA. Cả hai đều liên quan đến các enzyme tổng hợp một chuỗi nucleic acid mới bổ sung với khuôn mẫu DNA. Tất nhiên, hai quá trình này có những khác biệt quan trọng, mà đáng chú ý nhất là chuỗi mới trong quá trình phiên mã được tạo thành từ các ribonucleotide thay vì các deoxyribonucleotide. Các nguyên tắc cơ bản của quá trình phiên mã được thiết lập dựa vào nghiên cứu trên prokaryote (*E. coli*) nhưng dường như các nguyên tắc này cũng có tính phổ biến cho cả eukaryote. Tuy nhiên, do sự khác biệt về cấu trúc genome và hệ

thống enzyme nên sự phiên mã ở prokaryote và eukaryote cũng có những đặc trưng nhất định.

I. Các đặc điểm cơ bản của quá trình phiên mã

1. Sự phiên mã tạo ra RNA bổ sung với một sợi DNA

Các ribonucleotide nối tiếp của RNA được xác định dựa trên nguyên tắc bổ sung với các nucleotide trên DNA khuôn mẫu. Khi sự bắt cặp chính xác xảy ra, ribonucleotide tiếp theo được liên kết đồng hóa trị với chuỗi RNA đang hình thành nhờ phản ứng có enzyme xúc tác. Như vậy, sản phẩm phiên mã được kéo dài từng ribonucleotide một và bổ sung chính xác với sợi DNA được dùng làm khuôn mẫu.

Quá trình phiên mã dù có tính chính xác cao nhưng vẫn kém hơn nhiều so với quá trình tái bản DNA (tỷ lệ mắc lỗi là 1/10.000 nucleotide so với 1/10.000.000 trong tái bản). Đó là do sự thiếu một cơ chế sửa sai hữu hiệu, mặc dù trong quá trình tổng hợp RNA cũng có hai dạng sửa sai tồn tại. Tuy nhiên, vì các RNA được phiên mã không bao giờ được sao chép lại nên các sai sót xảy ra không ảnh hưởng đến việc truyền đạt thông tin cho thế hệ sau.

2. Sự phiên mã là một phản ứng enzyme

Những enzyme chịu trách nhiệm cho quá trình phiên mã ở cả tế bào prokaryote và eukaryote đều được gọi là RNA polymerase phụ thuộc DNA (DNA-dependent RNA polymerase), gọi tắt là RNA polymerase.

RNA polymerase xúc tác hình thành các cầu nối phospho-diester để nối các ribonucleotide thành một chuỗi thẳng. Enzyme dịch chuyển từng bước dọc theo DNA khuôn mẫu và kéo dài chuỗi RNA theo hướng từ 5'→3', nghĩa là các ribonucleotide được thêm vào đầu 3' của chuỗi RNA đang hình thành. Các cơ chất được sử dụng để tổng hợp RNA là ATP, GTP, CTP và UTP. Cũng giống như trong sự tái bản DNA, năng lượng cho phản ứng được cung cấp từ sự thủy phân các cầu nối giàu năng lượng của các cơ chất nói trên.

3. Sự phiên mã chỉ sao chép chọn lọc một số phần của genome và tạo ra nhiều bản sao

Sự lựa chọn vùng nào để phiên mã không phải xảy ra ngẫu nhiên. Mỗi vùng phiên mã điển hình bao gồm một hoặc nhiều gen, có những trình tự DNA đặc hiệu hướng dẫn khởi đầu và kết thúc phiên mã.

Đối với một vùng được chọn phiên mã, có thể có một đến hàng trăm thậm chí cả nghìn bản sao RNA được tạo ra. Sự tổng hợp phân tử RNA sau được bắt đầu trước khi phân tử RNA trước hoàn thành. Từ một gen đơn độc, trong vòng một giờ có thể tổng hợp được hơn một nghìn phân tử RNA (đối với eukaryote).

Sự lựa chọn vùng nào để phiên mã và mức độ phiên mã đều được điều hòa. Vì vậy, trong những tế bào khác nhau hoặc trong cùng một tế bào nhưng ở những thời điểm khác nhau sẽ có những nhóm gen khác nhau được phiên mã.

4. Chỉ một trong hai sợi đơn của phân tử DNA được dùng làm khuôn mẫu

Việc gắn của RNA polymerase vào promoter của gen sẽ quyết định việc lựa chọn sợi nào trong hai sợi đơn của DNA làm khuôn mẫu. Promoter, ngoài việc mang vị trí gắn RNA polymerase còn chứa đựng thông tin xác định sợi nào trong hai sợi đơn của DNA được phiên mã và xác định vị trí bắt đầu phiên mã.

5. Sự phiên mã được khởi phát không cần mồi

RNA polymerase có thể khởi đầu sự tổng hợp RNA trên khuôn mẫu DNA mà không cần mồi như DNA polymerase. Điều này đòi hỏi ribonucleotide đầu tiên được mang đến vị trí bắt đầu phiên mã phải được giữ ổn định trên DNA khuôn mẫu trong khi ribonucleotide thứ hai đang được đưa đến để xảy ra phản ứng trùng hợp.

II. Các giai đoạn của quá trình phiên mã

RNA polymerase tiến hành quá trình phiên mã một gen thông qua nhiều bước và được chia thành ba giai đoạn: khởi đầu, kéo dài, và kết thúc.

1. Giai đoạn khởi đầu

Đầu tiên, RNA polymerase gắn với promoter của gen (cùng với các yếu tố khởi đầu) để tạo thành phức hợp promoter-polymerase. Một khi được hình thành, phức hợp này sẽ có sự thay đổi cấu trúc cần thiết cho việc tiến hành giai đoạn khởi đầu.

Giai đoạn khởi đầu này bao gồm ba bước như sau:

1.1. Phức hợp đóng (closed complex)

Khi RNA polymerase vừa mới gắn với promoter thì phức hợp tạo thành ở trạng thái đóng. Ở trạng thái này, DNA vẫn là chuỗi kép và enzyme gắn vào một phía của vòng xoắn.

1.2. Phức hợp mở (open complex)

Lúc này DNA xung quanh điểm bắt đầu phiên mã được mở xoắn và liên kết giữa các cặp base bổ sung bị phá vỡ. Hai sợi DNA với độ dài khoảng 14 bp xung quanh vị trí bắt đầu phiên mã tách rời nhau ra và tạo nên vòng phiên mã. Việc mở DNA đã giải phóng sợi khuôn mẫu. Hai nucleotide đầu tiên được mang đến vị trí bắt đầu đã được hoạt hóa, chúng xếp hàng trên khuôn mẫu và được nối với nhau. RNA polymerase bắt đầu chuyển

dịch dọc theo sợi khuôn mẫu, mở xoắn phía trước vị trí trùng hợp và tái gắn hai sợi DNA phía sau. Bằng cách này, các ribonucleotide tiếp theo được gắn vào chuỗi RNA đang phát triển. Đoạn RNA ban đầu với số lượng ribonucleotide ít hơn 10 sẽ bị loại bỏ và enzyme bắt đầu tổng hợp lại.

1.3. Phức hợp tam nguyên bền (stable ternary complex)

Một khi enzyme đi được khoảng hơn 10 bp, nó được xem là "thoát" khỏi promoter. Lúc này một phức hợp gồm ba thành phần là enzyme, DNA và RNA được hình thành và quá trình phiên mã chuyển sang giai đoạn kéo dài.

2. Giai đoạn kéo dài

Một khi RNA polymerase tổng hợp được một đoạn RNA khoảng 10 base, nó chuyển sang giai đoạn kéo dài. Sự chuyển giai đoạn này đòi hỏi những biến đổi hơn nữa về cấu hình của RNA polymerase làm cho nó càng giữ chặt khuôn mẫu hơn nữa. Trong quá trình kéo dài, enzyme này thực hiện một loạt nhiều nhiệm vụ rất quan trọng ngoài việc tổng hợp RNA. Nó mở xoắn DNA trước mặt và tái gắn DNA phía sau, nó tách dần chuỗi RNA đang hình thành ra khỏi khuôn mẫu khi đang di chuyển dọc theo gen, và nó còn thực hiện chức năng đọc sửa sai (ở quá trình tái bản DNA, các chức năng này do nhiều enzyme khác nhau thực hiện).

3. Giai đoạn kết thúc

Khi RNA polymerase phiên mã hết chiều dài của gen (hoặc các gen), nó dừng lại và giải phóng sản phẩm RNA. Trong một số tế bào, có những trình tự đặc hiệu thúc đẩy giai đoạn kết thúc. Trong một số tế bào khác, người ta vẫn chưa rõ yếu tố nào điều khiển enzyme ngừng phiên mã và giải phóng RNA khỏi khuôn mẫu.

III. Phiên mã ở prokaryote

1. Enzyme RNA polymerase ở prokaryote

Tế bào prokaryote chỉ chứa một loại RNA polymerase. Enzyme này được cấu tạo bởi năm tiểu đơn vị, gồm hai tiểu đơn vị α và các tiểu đơn vị β, β', ω. Các tiểu đơn vị này liên kết với nhau để tạo thành enzyme lõi.

Người ta làm thí nghiệm tinh sạch enzyme lõi từ tế bào vi khuẩn rồi cho vào dung dịch chứa DNA vi khuẩn và ribonucleotide thì enzyme sẽ gắn với DNA và tổng hợp RNA. Tuy nhiên, RNA thu được trong thí nghiệm này không giống với RNA trong tế bào vì enzyme đã gắn vào các vị trí không thích hợp trên DNA. Nếu cho thêm một loại polypeptide tinh sạch được gọi là yếu tố σ trước khi enzyme gắn với DNA thì sự phiên mã sẽ bắt đầu tại những vị trí chính xác như trong tế bào.

Việc gắn yếu tố σ vào enzyme lõi làm tăng ái lực của enzyme với vị trí khởi động trên DNA, đồng thời làm giảm ái lực đối với các vùng khác. Khi enzyme lõi được gắn thêm yếu tố σ sẽ trở thành dạng holoenzyme.

2. Promoter của gen ở prokaryote

Promoter của gen vi khuẩn nằm ngay trước vị trí khởi đầu phiên mã. Bằng cách phân tích các trình tự DNA này của một số lượng lớn các gen vi khuẩn người ta nhận thấy có hai đoạn ngắn tương đối giống nhau giữa gen này và gen khác, mỗi đoạn gồm sáu nucleotide. Đoạn thứ nhất nằm cách vị trí khởi đầu khoảng 35 bp, là sự biến đổi của trình tự TTGACA. Đoạn thứ hai nằm cách vị trí khởi đầu khoảng 10 bp, là sự biến đổi của trình tự TATAAT. Chúng lần lượt được gọi là vùng –35 và vùng –10 (hộp Pribnow).

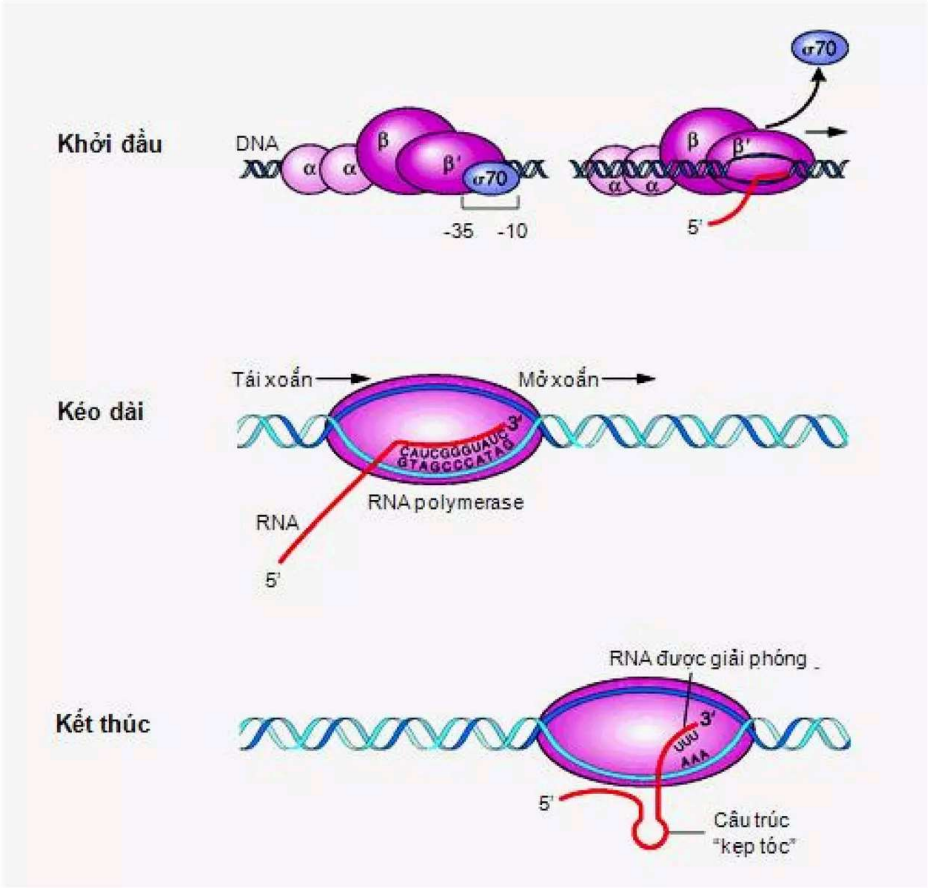
Đối với các promoter mạnh (ví dụ ở gen mã hóa rRNA), người ta còn tìm thấy yếu tố UP làm tăng sự gắn của RNA polymerase vào DNA. Có một số promoter thiếu vùng –35 và được thay bằng yếu tố –10 mở rộng, bao gồm vùng –10 chuẩn và thêm một đoạn ngắn ở đầu 5’ của nó (ví dụ ở gen *gal* của *E. coli*).

Yếu tố σ nhận dạng các vùng –35 và –10 hoặc yếu tố –10 mở rộng nhờ các cấu trúc đặc biệt của chúng. Riêng yếu tố UP không được nhận dạng bởi σ mà được nhận dạng bởi một vùng ở đầu C tận cùng của tiểu đơn vị α, gọi là α CTD (carboxyl terminal domain).

3. Vai trò của enzyme RNA polymerase và promoter trong quá trình phiên mã

Khi khởi đầu sự phiên mã, RNA polymerase gắn với DNA ở vùng –35 một cách lỏng lẻo tạo thành phức hợp đóng. Sau đó, enzyme gắn vào promoter chặt hơn và phức hợp chuyển sang trạng thái mở. Một vùng DNA được mở xoắn nằm giữa hai vị trí –11 và +3, nghĩa là bao gồm cả vị trí bắt đầu phiên mã +1. Sợi đơn DNA làm khuôn mẫu được bộc lộ để sinh tổng hợp RNA.

Một khi RNA polymerase tổng hợp được 10 nucleotide (sau lần khởi đầu) thì yếu tố σ rời khỏi enzyme và có thể đến gắn vào promoter khác để khởi đầu một quá trình phiên mã mới. Sự tách rời yếu tố σ cho phép hình thành một kênh thoát mà thông qua đó phần RNA đã được tổng hợp chui ra khỏi enzyme và được kéo dài (Hình 5.1).



Hình 5.1. Các giai đoạn phiên mã ở prokaryote

4. Tín hiệu kết thúc

Tín hiệu kết thúc là những trình tự trên DNA có vai trò thúc đẩy sự tách RNA polymerase ra khỏi DNA và giải phóng chuỗi RNA mới được tổng hợp. Ở vi khuẩn, có hai loại tín hiệu kết thúc là tín hiệu kết thúc không phụ thuộc Rho và tín hiệu kết thúc phụ thuộc Rho.

4.1. Tín hiệu kết thúc không phụ thuộc Rho

Đó là một cấu trúc đặc biệt bao gồm hai trình tự đối xứng bổ sung nhau giàu GC, tiếp theo sau là một loạt 8 adenine (chúng sẽ được phiên mã thành 8 uracil). Những trình tự này không ảnh hưởng đến RNA polymerase cho đến sau khi chúng được phiên mã. Khi đó các trình tự đối xứng bổ sung trên RNA sẽ bắt cặp với nhau và tạo nên cấu trúc hình chiếc kẹp tóc (Hình 5.2). Cấu trúc kẹp tóc này đã phá vỡ phức hợp kéo dài hoặc bằng cách thúc đẩy mở kênh thoát cho RNA trên RNA polymerase hoặc bằng cách phá vỡ mối tương tác RNA-khuôn mẫu DNA

Cấu trúc kẹp tóc này chỉ hoạt động hiệu quả khi nó được theo sau bởi một loạt U. Tại thời điểm cấu trúc kẹp tóc được hình thành, chuỗi RNA đang phát triển vẫn được giữ trên khuôn mẫu chỉ bằng các liên kết giữa A-U. Vì liên kết A-U yếu hơn G-C nên chúng dễ dàng bị phá vỡ bởi hiệu quả của cấu trúc kẹp tóc, và vì thế RNA dễ dàng được giải phóng.

4.2. Tín hiệu kết thúc phụ thuộc Rho

Tín hiệu này là những thành phần RNA mà muốn hoạt động, đòi hỏi phải có yếu tố Rho. Yếu tố Rho là một protein hình nhẫn (ở *E. coli* có khối lượng phân tử khoảng 50 kDa) gồm sáu tiểu đơn vị giống hệt nhau. Yếu tố này có hoạt tính ATPase: khi gắn vào sản phẩm phiên mã, nó sẽ sử dụng năng lượng từ sự thủy phân ATP để kéo RNA ra khỏi khuôn mẫu và polymerase.

Yếu tố Rho điều khiển phân tử RNA đặc biệt nhờ tính đặc hiệu. Thứ nhất là tính đặc hiệu tại những vị trí nó gắn, đó là những đoạn khoảng 40 nucleotide mà không gấp lại thành cấu trúc bậc hai, chúng thường giàu C. Tính đặc hiệu thứ hai là yếu tố Rho giảm gắn với bất kỳ sản phẩm phiên mã nào đang được dịch mã (nghĩa là những sản phẩm đang gắn ribosome). Ở vi khuẩn, phiên mã và dịch mã được song hành chặt chẽ: dịch mã khởi phát ngay trên RNA đang được tổng hợp kể từ khi RNA bắt đầu chui ra khỏi RNA polymerase. Vì vậy Rho chỉ kết thúc những sản phẩm vẫn đang được phiên mã ở quá đầu tận cùng của gen hoặc operon.

IV. Quá trình phiên mã ở eukaryote

1. Cấu trúc gen ở eukaryote

Cấu trúc một gen mã hóa cho protein của eukaryote bao gồm các vùng sau:

1.1. Vùng 5’ kiểm soát biểu hiện gen

Vùng này bao gồm các trình tự nucleotide điều hòa biểu hiện gen và hoạt hóa sự phiên mã, bao gồm:

- Promoter. Là những trình tự định vị ở đầu 5' không dịch mã của gen có chức năng xác định vị trí bắt đầu phiên mã, kiểm soát số lượng mRNA và đôi khi cả tính đặc hiệu mô (tissue-specific). Promoter có thể dài đến vài nghìn bp. Vùng này thường chứa một trình tự bảo thủ gọi là hộp TATA nằm cách vị trí phiên mã khoảng 25-30 bp. Hộp này giúp xác định chính xác vị trí bắt đầu phiên mã. Ngoài ra, còn có hộp CCAAT nằm cách vị trí bắt đầu phiên mã khoảng 75-80 bp. Hộp này ít phổ biến hơn hộp TATA, nó có chức năng tăng hiệu quả phiên mã. Một số gen "quản gia" mã hóa cho các enzyme hiện diện ở tất cả các tế bào thường thiếu cả hai hộp này và promoter rất giàu GC. Ngoài ra, còn có các thành phần đặc hiệu khác.

- Vị trí gắn vùng đặc hiệu mô. Là trình tự DNA tương tác với protein đặc hiệu mô để chỉ huy gen cấu trúc sản xuất ra từng protein đặc hiệu của từng mô.

- Vị trí gắn với vùng tăng cường phiên mã (enhancer). Còn gọi là gen tăng cường, là những trình tự DNA được gắn với các tác nhân hoạt hóa để kích thích phiên mã của các gen kề bên, chúng có thể tác động qua một khoảng cách xa và có thể tác động theo hai phía (từ 5' hoặc 3' tới).

1.2. Vùng được phiên mã

Bao gồm các exon và intron nằm xen kẽ. Đây là một đặc điểm để phân biệt với gen của prokaryote. Các exon và intron đều được phiên mã nhưng chỉ có các exon là được dịch mã. Các intron được bắt đầu bằng GT và kết thúc bằng AG. Các intron chiếm phần lớn trong mỗi gen và chúng sẽ được loại bỏ khỏi RNA mới được tổng hợp, còn các exon được nối với nhau để tạo nên mRNA hoàn chỉnh (mature mRNA).

Ở hai đầu của vùng được phiên mã (coding region) còn có vùng 5' không dịch mã (5' untranslation region) và vùng 3' không dịch mã (3' untranslation region). Vùng 5' không dịch mã được tính từ vị trí bắt đầu phiên mã cho đến codon khởi đầu ATG. Vùng 3' không dịch mã bắt đầu từ codon kết thúc đến vị trí gắn đuôi poly(A).

1.3. Vùng 3' không dịch mã

Chức năng chưa rõ, ở một số gen vùng này mang các trình tự điều hòa chuyên biệt.

2. Enzyme RNA polymerase của eukaryote

Tế bào eukaryote có đến ba loại RNA polymerase là RNA polymerase I (pol I), RNA polymerase II (pol II), và RNA polymerase III (pol III). Trong đó, RNA polymerase II đảm nhận việc phiên mã cho hầu hết các gen, chủ yếu là gen mã hóa cho protein. RNA polymerase I phiên mã các gen tiền thân của rRNA lớn và RNA polymerase III phiên mã các gen tRNA, một số gen RNA kích thước nhỏ của nhân (small nuclear, snRNA) và RNA 5S. Các RNA polymerase của eukaryote cũng bao gồm nhiều

tiểu đơn vị, trong đó có những tiểu đơn vị tương ứng với các tiểu đơn vị trong enzyme của vi khuẩn.

Bảng 5.1. Các tiểu đơn vị của RNA polymerase

Prokaryote		Eukaryote		
Vi khuẩn	Archaea	RNA pol I	RNA pol II	RNA pol III
β'	A'/A"	RPA1	RPB1	RPC1
β	B	RPA2	RPB2	RPC2
α'	D	RPC5	RPB3	RPC5
α"	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	(+6)	(+9)	(+7)	(+11)

3. Các yếu tố giúp RNA polymerase khởi đầu phiên mã

Trong khi RNA polymerase của prokaryote chỉ cần thêm một yếu tố khởi đầu là σ để khởi động phiên mã thì RNA polymerase của eukaryote phải cần nhiều yếu tố khởi đầu, các yếu tố này được gọi là các yếu tố phiên mã tổng quát. Mặt khác, sự khởi đầu phiên mã ở eukaryote còn phải đối mặt với hiện tượng các DNA được đóng gói trong nucleosome và các dạng cao hơn của cấu trúc chromatin. Điều này đòi hỏi phải có nhiều yếu tố khác thêm vào để giúp khởi động quá trình phiên mã.

3.1. Vai trò của các yếu tố phiên mã tổng quát

Các yếu tố phiên mã tổng quát giúp RNA polymerase vào đúng vị trí trên promoter, giúp tách hai sợi đơn của DNA ra để phiên mã được bắt đầu, và giải phóng RNA polymerase khỏi promoter một khi phiên mã đã được khởi động xong để đi vào giai đoạn kéo dài.

Các yếu tố này được gọi là "tổng quát" vì chúng gắn trên tất cả các promoter được sử dụng bởi RNA polymerase II. Do đó, chúng được viết tắt là TFII (transcription factor for polymerase II), bao gồm TFIIA, TFIIB...

Như đã mô tả ở trên, nhiều promoter của eukaryote chứa hộp TATA. Hộp này được nhận dạng bởi một tiểu đơn vị của TFIID là TBP (TATA binding protein: protein gắn TATA). Phức hợp TBP-DNA tạo nên một cái nền để thu hút các TFII khác và RNA polymerase đến promoter. *In vitro*, các yếu tố phiên mã tổng quát khác đến gắn vào promoter theo thứ tự sau: TFIIA, TFIIB, TFIIF cùng RNA polymerase II, TFIIE và TFIIH. Sau đó, vùng promoter được mở xoắn. Khác với vi khuẩn, sự mở xoắn ở đây cần có năng lượng cung cấp từ sự thủy phân ATP nhờ TFIIH (yếu tố này có hoạt tính giống helicase).

Sau đó cũng xảy ra hiện tượng khởi đầu sẩy giống như ở prokaryote cho đến khi có sự biến đổi về cấu hình nhằm giải phóng RNA polymerase II khỏi promoter và đi vào giai đoạn kéo dài. Tuy nhiên, ở đây có một bước mà không tìm thấy ở prokaryote đó là sự gắn thêm các gốc phosphate vào đuôi của RNA polymerase (đuôi này còn được gọi là CTD: carboxyl terminal domain). Sự phosphoryl hóa này cũng được xúc tác bởi TFIIH nhờ hoạt tính protein kinase của nó (Hình 5.3).

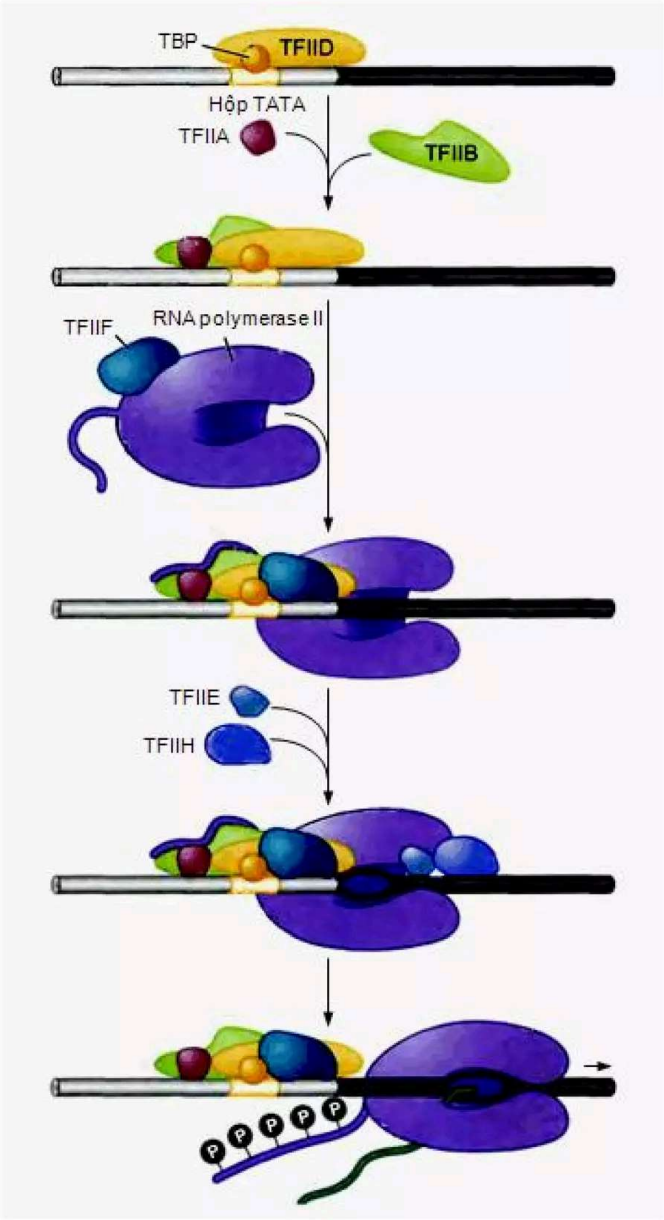
3.2. Vai trò của các tác nhân hoạt hóa, phức hợp trung gian và enzyme biến đổi chromatin.

Ngoài các yếu tố phiên mã tổng quát, RNA polymerase II còn cần sự hỗ trợ của các yếu tố khác (Hình 5.4).

Đầu tiên, cần có các tác nhân hoạt hóa phiên mã (transcriptional activator) đến gắn vào các trình tự đặc hiệu trên DNA (các enhancer) để hấp dẫn RNA polymerase II đến vị trí bắt đầu phiên mã. Sự hấp dẫn này cần thiết để RNA polymerase II và các yếu tố phiên mã tổng quát vượt qua trở ngại khi gắn với DNA được đóng gói trong chromatin.

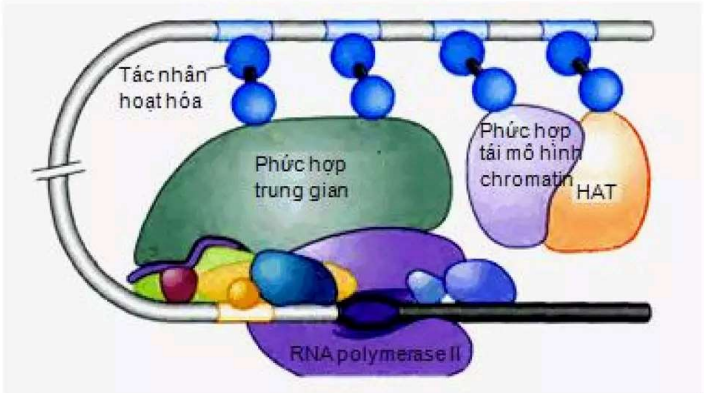
Tiếp đến, sự khởi đầu phiên mã *in vivo* còn cần sự hiện diện của các protein tạo thành phức hợp trung gian (mediator complex). Phức hợp này cho phép tác nhân hoạt hóa tác động tốt lên RNA polymerase II và các yếu tố phiên mã tổng quát.

Cuối cùng, sự phiên mã còn cần các enzyme biến đổi chromatin, bao gồm phức hợp tái tạo mô hình chromatin (chromatin remodeling complex) và enzyme histone acetylase. Cả hai có tác dụng giúp cho bộ máy khởi đầu phiên mã có thể gắn vào DNA trong chromatin một cách dễ dàng.



Hình 5.3. Các yếu tố phiên mã tổng quát giúp khởi đầu phiên mã

Như vậy, có rất nhiều protein gắn vào promoter để khởi động sự phiên mã ở eukaryote. Thứ tự gắn của các protein này thay đổi đối với các gen khác nhau. Thực tế, một số có thể gắn với nhau ở xa DNA rồi được mang đến DNA dưới dạng phức hợp. Ví dụ: phức hợp trung gian, RNA polymerase II, và một số yếu tố phiên mã tổng quát có thể gắn với nhau trong nhân tương rồi được mang đến DNA.



Hình 5.4. Các tác nhân hoạt hóa, phức hợp trung gian và các thành phần biến đổi nucleosome trong phiên mã

4. Các yếu tố kích thích RNA polymerase II hoạt động trong giai đoạn kéo dài

Một khi RNA polymerase II bắt đầu chuyển sang giai đoạn kéo dài thì các yếu tố khởi đầu được loại bỏ như yếu tố phiên mã tổng quát và phức hợp trung gian. Thay vào đó, các yếu tố kích thích giai đoạn kéo dài được thu hút đến, bao gồm TFIIS và hSPT5. Ngoài ra, còn có các yếu tố khác được huy động để phục vụ cho quá trình biến đổi RNA mới được tổng hợp.

Giai đoạn kéo dài trong phiên mã được song hành chặt chẽ với các bước biến đổi RNA mới được tổng hợp. Như đã phân tích ở trên, có một bước quan trọng diễn ra khi chuyển từ giai đoạn khởi đầu sang kéo dài là sự phosphoryl hóa đuôi CTD của RNA polymerase II. Sự phosphoryl hóa này không những giúp RNA polymerase II thoát khỏi các protein ở vị trí bắt đầu phiên mã mà còn cho phép các protein khác đến kết hợp với đuôi để giúp quá trình kéo dài phân tử RNA và biến đổi RNA tiền thân.

5. Quá trình biến đổi các RNA mới được tổng hợp

Ở eukaryote, sự phiên mã chỉ là bước đầu tiên trong một loạt các phản ứng bao gồm cả sự biến đổi hai đầu của mRNA tiền thân và loại bỏ các intron để tạo thành mRNA hoàn chỉnh.

5.1. Sự gắn mũ vào đầu 5’

Ngay khi RNA polymerase II vừa mới tạo ra khoảng 25 ribonucleotide của RNA, đầu 5’ của phân tử RNA này được biến đổi bằng cách gắn thêm một cái mũ là guanine có biến đổi hóa học. Phản ứng gắn mũ được thực hiện bởi ba loại enzyme là:

- Phosphatase có tác dụng loại một gốc phosphate khỏi đầu 5’ của RNA mới sinh.

- Guanylyl transferase gắn GMP bằng liên kết đảo ngược (5’ với 5’ thay vì 5’ với 3’) vào đầu 5’ của RNA đang được tổng hợp.

- Methyl transferase gắn nhóm methyl vào guanosine (Hình 5.5).

Trong nhân, mũ gắn với một phức hợp protein được gọi là CBC (CAP-binding complex: phức hợp gắn mũ) giúp RNA được xử lý tốt và được chuyển ra ngoài. Mũ 5' còn có vai trò quan trọng trong quá trình dịch mã mRNA trong bào tương.

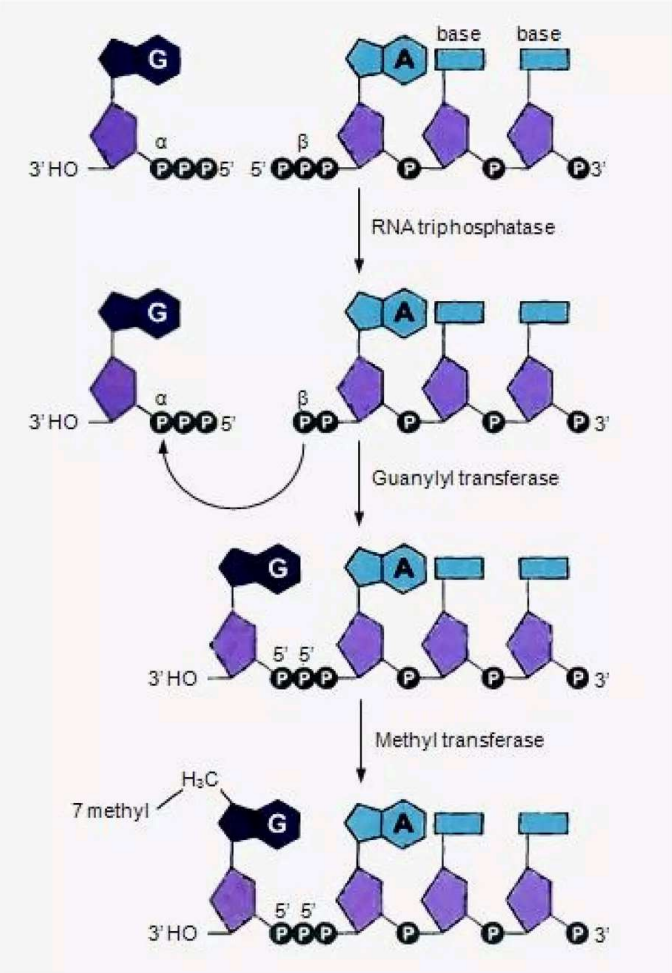
5.2. Sự gắn đuôi poly(A) vào đầu 3' và kết thúc phiên mã

Sự gắn đuôi poly(A) vào đầu 3' được liên kết với sự kết thúc phiên mã. Các enzyme cần thiết cho các quá trình này được tập trung trên đuôi CTD của RNA polymerase II, bao gồm CPSF (cleavage and polyadenylation specificity factor: yếu tố đặc hiệu tách RNA và gắn đuôi polyA) và CstF (cleavage stimulation factor: yếu tố kích thích tách RNA).

Khi RNA polymerase di chuyển đến cuối gen, nó gặp những trình tự đặc hiệu (được gọi là tín hiệu polyA). Trình tự này được phiên mã thành RNA, thúc đẩy chuyển CPSF và CstF đến gắn vào đoạn RNA này. Sau đó các protein khác sẽ được tập trung đến để khởi đầu sự tách rời RNA và gắn đuôi poly(A).

Sự gắn đuôi được thực hiện bởi enzyme poly(A) polymerase (PAP). Enzyme này thêm khoảng từ vài chục đến 200-250 adenine vào đầu 3' của RNA đã được tách ra. Enzyme PAP sử dụng ATP làm cơ chất và hoạt động giống RNA polymerase, tuy nhiên không có khuôn mẫu. Người ta vẫn chưa rõ yếu tố nào quyết định chiều dài của đuôi nhưng quá trình này có liên quan đến các protein gắn đặc hiệu với đuôi poly(A).

Sau khi RNA được tách ra và gắn đuôi poly(A), RNA polymerase vẫn chưa kết thúc phiên mã ngay. Nó còn tiếp tục di chuyển dọc theo khuôn mẫu và tạo ra một phân tử RNA thứ hai có thể dài hàng trăm nucleotide trước khi kết thúc. Sau đó, RNA polymerase được tách ra khỏi khuôn mẫu, giải phóng RNA mới và RNA này sẽ bị giáng hóa.



Hình 5.5. Phản ứng gắn mũ vào đầu 5' của RNA

5.3. Quá trình cắt nối gen (splicing)

Đây là quá trình loại bỏ các intron và nối các exon lại với nhau. Quá trình này được thực hiện phần lớn bởi các RNA thay vì protein.

Các phân tử RNA này tương đối ngắn (dưới 200 nucleotide) được gọi là snRNA. Có năm loại liên quan đến dạng cắt nối chính là U1, U2, U4, U5 và U6. Mỗi snRNA được kết hợp với nhiều protein để hình thành snRNP.

Có ba trình tự nằm trên intron đóng vai trò quan trọng trong quá trình cắt nối là: vị trí cắt nối đầu 5', vị trí phân nhánh là một trình tự giàu các pyrimidine bao quanh một nucleotide adenine ở gần đầu 3', và vị trí cắt nối đầu 3'.

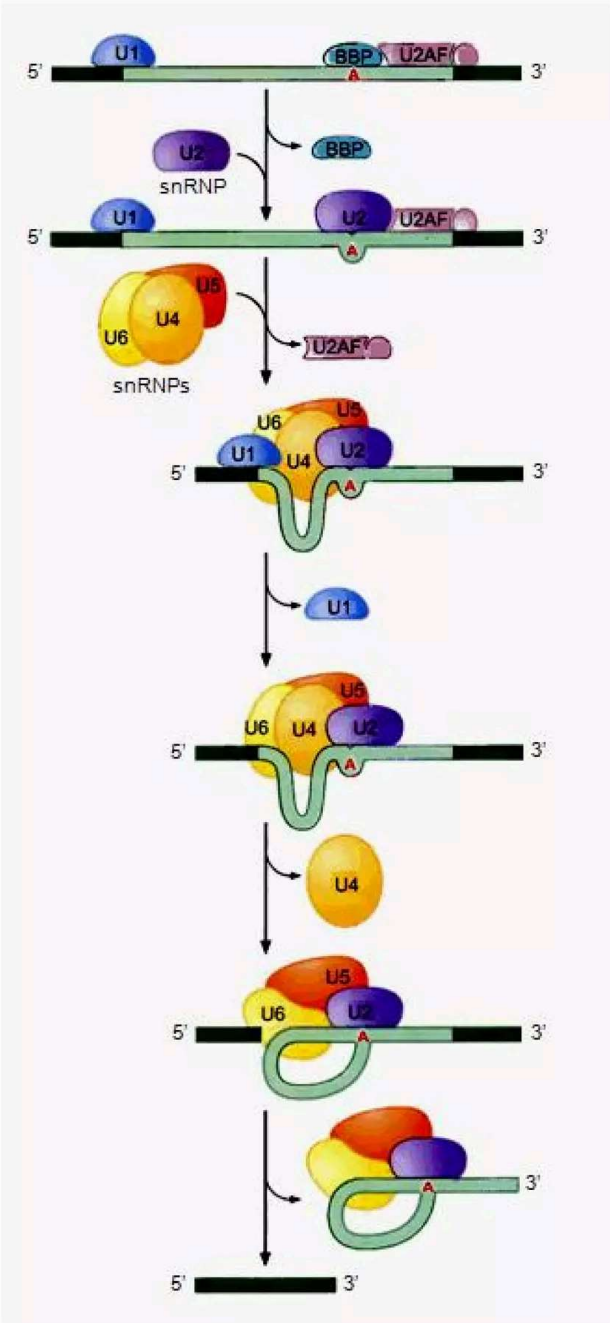
Đầu tiên, vị trí cắt nối đầu 5' được nhận dạng bởi U1 snRNP bằng sự bắt cặp các base bổ sung. Vị trí phân nhánh cũng được nhận dạng bởi BBP (branch-point binding protein: protein gắn vị trí phân nhánh) và U2AF. U2 snRNP đến thay thế BBP (nhờ

U2AF giúp đỡ). Sự bắt cặp base giữa U2 snRNA và vị trí phân nhánh làm thừa ra một gốc A không bắt cặp và sẵn sàng phản ứng với vị trí 5'. Sau đó U4 và U6 snRNP cùng U5 snRNP đến gắn với phức hợp trên.

Dạng U1 rời khỏi phức hợp và U6 thế chỗ U1 tại vị trí cắt nối đầu 5'. U4 được giải phóng để U6 tương tác với U2. Điều này làm cho vị trí cắt nối đầu 5' và vị trí phân nhánh được đặt kề nhau và tạo điều kiện cho phản ứng cắt nối xảy ra. Nucleotide adenine đặc hiệu ở vị trí phân nhánh tấn công và cắt intron ở vị trí đầu 5'. Đồng thời có một liên kết đồng hóa trị xảy ra giữa A ở vị trí phân nhánh và đầu 5' của intron tạo nên một cấu trúc hình thòng lọng. Đầu 3' của exon trước nối với đầu 5' của exon kế tiếp và thòng lọng intron được giải phóng (Hình 5.6).

V. Phiên mã ngược

Phiên mã ngược là quá trình tổng hợp DNA dựa trên khuôn mẫu RNA. Quá trình này được xúc tác bởi một loại enzyme đặc biệt gọi là enzyme phiên mã ngược (RT: reverse transcriptase). Enzyme này có cả hoạt tính DNA polymerase và hoạt tính RNase H. Cũng giống như DNA polymerase trong quá trình tái bản DNA, enzyme phiên mã ngược đòi hỏi phải có primer (mồi) đặc biệt để tổng hợp nên DNA mới.



Hình 5.6. Quá trình cắt nối gen

Enzyme phiên mã ngược được phát hiện lần đầu tiên ở retrovirus. Sau khi các retrovirus đi vào tế bào vật chủ, genome của RNA của nó sẽ được phiên mã ngược thành

DNA sợi đôi, rồi tích hợp vào DNA của vật chủ. Quá trình phiên mã ngược này khá phức tạp, người ta có thể tóm tắt thành 10 bước như sau:

- Bước 1: Một tRNA của tế bào có tính đặc hiệu với retrovirus đóng vai trò primer để khởi phát quá trình phiên mã ngược. Primer lai với vùng bổ sung trên genome của RNA của retrovirus gọi là vị trí gắn primer (PBS: primer-binding site).

- Bước 2: Một đoạn DNA được kéo dài từ tRNA có trình tự bổ sung với trình tự của genome RNA của retrovirus.

- Bước 3: Trình tự R đầu 5' và U5 của virus bị loại bỏ bởi RNase H.

- Bước 4: Đổi khuôn mẫu lần thứ nhất (first template exchange) còn được gọi là "nhảy" lần một: DNA đến lai với trình tự R còn lại ở đầu 3'.

- Bước 5: Một sợi DNA được kéo dài từ đầu 3'.

- Bước 6: Hầu hết RNA virus bị loại bỏ bởi RNase H.

- Bước 7: Sợi DNA thứ hai được kéo dài từ RNA còn lại của virus.

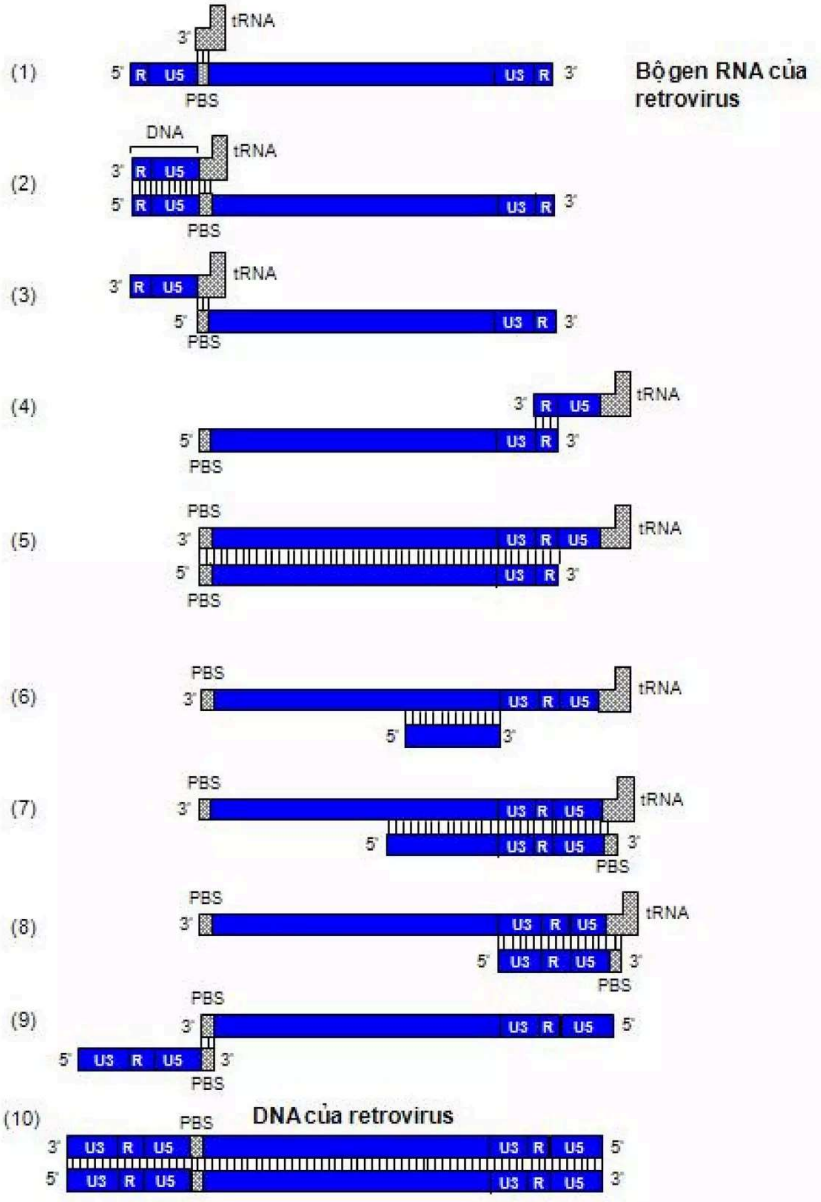
- Bước 8: Cả tRNA và phần RNA còn lại của virus bị loại bỏ bởi RNase H.

- Bước 9: Đổi khuôn mẫu lần thứ hai ("nhảy" lần hai): vùng PBS của sợi DNA thứ hai đến lai với vùng PBS của sợi thứ nhất.

- Bước 10: Kéo dài cả hai sợi DNA.

R là những trình tự lặp lại trên genome của retrovirus (ở đầu 5' và 3'), U5 và U3 là những vùng mã hóa cho những tín hiệu tích hợp ở đầu 5' và 3' (Hình 5.7).

Ngày nay, người ta cũng phát hiện enzyme phiên mã ngược ở các virus động vật khác như hepadnavirus, và ở các virus thực vật như caulimovirus. Tất cả chúng được gọi là retrovirus. Ngoài ra, người ta còn phát hiện hoạt tính enzyme phiên mã ngược ở một số dòng của myxobacteria và *E. coli*. Enzyme phiên mã ngược đã trở thành một công cụ không thể thiếu trong sinh học phân tử. Nó giúp các nghiên cứu viên phiên mã ngược mRNA của tế bào thành cDNA (complementary DNA), rồi sau đó có thể khuếch đại, tạo dòng và biểu hiện bằng các phương pháp đặc biệt. Sự phát hiện enzyme phiên mã ngược ở nhiều loại virus và đặc biệt là ở một số vi khuẩn cũng có ý nghĩa quan trọng trong việc nghiên cứu sự tiến hóa của hệ thống sinh giới.



Hình 5.7. Quá trình phiên mã ngược của retrovirus

Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. *NXB Giáo dục*, Hà Nội.

2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. 2002. Molecular Biology of the Cell. 4th ed. *Garland Science Publishing, Inc.* New York, USA.

3. Flint SJ. 2000. Principles of Virology: Molecular Biology, Pathogenesis, and Control. *ASM Press*, Washington, USA.

4. Gelehrter TD, Collins FS and Ginsburg D. 1998. Principles of Medical Genetics. 2nd ed. *Williams & Wilkins Company,* Baltimore, USA.

5. Griffiths AJF, Wessler SR, Lewontin RC, Gelbart WM, Suzuki DT and Miller JH. 2005. An Introduction to Genetic Analysis. *WH Freeman & Co.* New York, USA.

6. Karp G. 2002. Cell and Molecular Biology. *John Wiley & Sons, Inc.* New York, USA.

7. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Losick R. 2004. Molecular Biology of the Gene. *Pearson Education, Inc./Benjamin Cummings Publishing*, San Francisco, USA.

Chương 6

Dịch mã

Dịch mã là quá trình các thông tin di truyền chứa trong các trình tự nucleotide của mRNA được sử dụng để tạo ra các chuỗi amino acid trong protein. Sự tổng hợp một protein riêng lẻ đòi hỏi sự tham gia của hơn 100 protein và RNA. Bộ máy dịch mã bao gồm bốn thành phần quan trọng là mRNA, tRNA, aminoacyl tRNA synthetase và ribosome. Các mRNA là khuôn mẫu cho quá trình dịch mã. Dịch mã là một trong những quá trình có tính bảo thủ cao và chiếm nhiều năng lượng của tế bào. Tuy nhiên, do cấu trúc khác nhau giữa mRNA của prokaryote và eukaryote nên quá trình dịch mã của chúng cũng có những điểm khác biệt quan trọng.

I. Mã di truyền

1. Các codon

Do chỉ có bốn loại nucleotide khác nhau trong mRNA và có đến 20 loại amino acid trong protein nên sự dịch mã không thể được thực hiện theo kiểu tương ứng một nucleotide-một amino acid được. Chuỗi nucleotide của một gen thông qua trung gian mRNA được dịch mã thành chuỗi amino acid của protein theo những quy luật được gọi là mã di truyền.

Người ta đã giải mã toàn bộ các amino acid vào những năm đầu của thập kỷ 1960. Mỗi amino acid được mã hóa bởi ba nucleotide liên tiếp trên DNA (hoặc RNA tương ứng), bộ ba nucleotide này được gọi là một codon. Với 4 loại nucleotide khác nhau sẽ có $4^3 = 64$ codon khác nhau được phân biệt bởi thành phần và trật tự của các nucleotide. Trong số này có 3 codon kết thúc (stop codon) là UAA, UAG và UGA có nhiệm vụ báo hiệu chấm dứt việc tổng hợp chuỗi polypeptide. Trong 61 mã còn lại có nhiều codon cùng mã hóa cho một amino acid (Bảng 3.4-Chương 3).

Mã di truyền có tính đồng nhất cho toàn bộ sinh giới trừ một số ngoại lệ đối với các codon ở ty thể. Ở DNA của bào quan này có một số codon mã hóa cho các amino acid khác với nghĩa của các codon này trên DNA trong nhân. Ví dụ:

- UGA mã hóa cho tryptophan thay vì báo hiệu chấm dứt việc tổng hợp protein.

- AGA và AGG không mã hóa cho arginine mà báo hiệu chấm dứt tổng hợp protein.

- AUA mã hóa cho methionine thay vì mã hóa cho isoleucine.

2. Các quy tắc chi phối mã di truyền

Có ba quy tắc điều khiển sự sắp xếp và sử dụng các codon trên mRNA là:

- Các codon được đọc theo hướng 5'→3'. Vì vậy chuỗi mã hóa cho dipeptide NH_2-Thr-Arg-COOH được viết là 5'-ACGCGA-3'.

- Các codon không chồng lên nhau và vùng dịch mã của mRNA không chứa các khoảng trống.

- Thông tin được dịch mã theo một khung đọc (reading frame) cố định. Về mặt nguyên tắc, cùng một trình tự RNA có thể có ba khung đọc khác nhau. Tuy nhiên, trên thực tế chỉ có một trong ba khung đọc này chứa thông tin thực sự, chính codon khởi đầu đã xác định khung đọc đúng cho mỗi trình tự mRNA.

II. Các ribosome

Ribosome là bộ máy đại phân tử điều khiển sự tổng hợp protein. Nó được cấu tạo bởi ít nhất là 3 phân tử RNA và trên 50 protein khác nhau[1] với khối lượng phân tử là 2,5 MDa (megadalton) đối với ribosome của prokaryote và 4,2 MDa đối với ribosome của eukaryote.

1. Thành phần cấu tạo của ribosome

Mỗi ribosome bao gồm một tiểu đơn vị lớn và một tiểu đơn vị nhỏ. Tiểu đơn vị lớn chứa trung tâm peptidyl transferase chịu trách nhiệm cho việc hình thành các cầu nối peptide. Tiểu đơn vị nhỏ chứa trung tâm giải mã, là nơi các tRNA đã được gắn amino acid đọc và giải mã các codon. Ngoài ra còn có trung tâm gắn các yếu tố ở tiểu đơn vị lớn.

Theo quy ước, các tiểu đơn vị được đặt tên theo tốc độ lắng của chúng dưới lực ly tâm. Đơn vị đo tốc độ lắng là Svedberg (tên của nhà phát minh máy siêu ly tâm) và được viết tắt là S. Ribosome của prokaryote là ribosome 70S, trong đó tiểu đơn vị lớn là 50S và tiểu đơn vị nhỏ là 30S. Ribosome của eukaryote là 80S, với tiểu đơn vị lớn là 60S và tiểu đơn vị nhỏ là 40S.

Mỗi tiểu đơn vị đều được cấu tạo bởi các RNA ribosome (rRNA) và các protein ribosome. Đơn vị Svedberg lại được sử dụng để phân biệt các rRNA (Bảng 6.1).

Trong quá trình dịch mã, tiểu đơn vị lớn và tiểu đơn vị nhỏ của mỗi ribosome liên kết với nhau và với mRNA. Sau mỗi vòng tổng hợp protein, chúng lại rời nhau ra.

Bảng 6.1. Các thành phần cấu tạo của ribosome

Các thành phần cấu tạo	Prokaryote		Eukaryote	
	Tiểu đơn vị lớn (50S)	Tiểu đơn vị nhỏ (30S)	Tiểu đơn vị lớn (60S)	Tiểu đơn vị nhỏ (40S)
rRNA	5S rRNA (120 Nu) 23S rRNA (2900 Nu)	16S rRNA (1540 Nu)	5,8S rRNA (160 Nu) 5S rRNA (120 Nu) 28S rRNA (4700 Nu)	18S rRNA (1900 Nu)
Protein	34 protein	21 protein	49 protein	33 protein

2. Khái niệm polyribosome

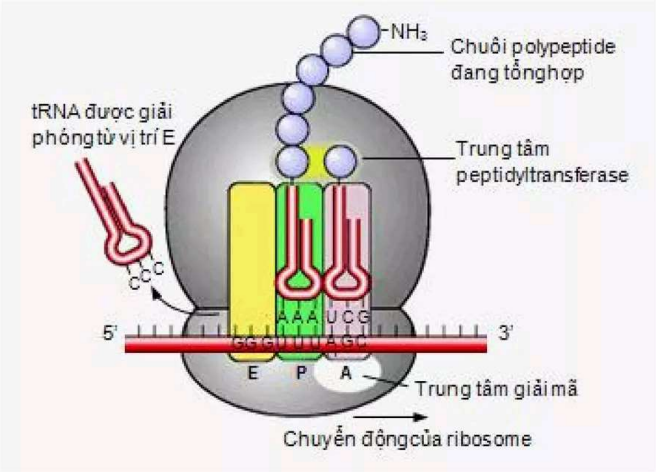
Mặc dù một ribosome chỉ có thể tổng hợp một polypeptide tại một thời điểm, nhưng mỗi mRNA có thể được dịch mã đồng thời bởi nhiều ribosome. Một mRNA mang nhiều ribosome được xem là polyribosome hay polysome. Mỗi ribosome đơn độc tiếp xúc với khoảng 30 nucleotide, nhưng do kích thước lớn của ribosome nên mật độ cho phép trên mRNA là 80 nucleotide cho mỗi ribosome.

3. Các vị trí gắn tRNA trên ribosome

Trên ribosome chứa ba vị trí gắn tRNA là vị trí A, P và E. Trong đó:

- A là vị trí gắn aminoacyl-tRNA (tRNA có mang amino acid).
- P là vị trí gắn peptidyl-tRNA (tRNA có mang chuỗi polypeptide).
- E (exit) là vị trí gắn tRNA mà được phóng thích sau khi chuỗi polypeptide được chuyển sang aminoacyl-tRNA.

Mỗi vị trí gắn tRNA được hình thành tại giao diện giữa tiểu đơn vị lớn và tiểu đơn vị nhỏ. Bằng cách này, các tRNA được gắn vào có thể bắt ngang qua khoảng cách giữa trung tâm peptidyl transferase của tiểu đơn vị lớn và trung tâm giải mã của tiểu đơn vị nhỏ. Đầu 3' của tRNA được nằm gần tiểu đơn vị lớn và vòng đối mã gần tiểu đơn vị nhỏ.



Hình 6.1. Các thành phần chức năng của ribosome

4. Các kênh của ribosome

Đó là các kênh cho phép mRNA đi vào và đi ra khỏi ribosome, và kênh cho phép chuỗi polypeptide mới sinh đi ra khỏi ribosome.

mRNA đi vào và đi ra khỏi trung tâm giải mã của ribosome thông qua hai kênh hẹp tại tiểu đơn vị nhỏ. Trong đó, kênh vào có chiều rộng chỉ đủ cho RNA không bắt cặp đi qua. Đặc điểm này đảm bảo cho mRNA được duỗi thẳng khi nó đi vào trung tâm giải mã, bằng cách loại bỏ mọi tương tác bắt cặp base bổ sung nội phân tử.

Một kênh xuyên qua tiểu đơn vị lớn tạo lối thoát cho chuỗi polypeptide mới được tổng hợp. Kích thước của kênh đã hạn chế được sự gấp của các chuỗi polypeptide đang tổng hợp. Vì vậy, protein chỉ có thể hình thành cấu trúc bậc ba sau khi nó được giải phóng khỏi ribosome.

III. Sự hình thành aminoacyl-tRNA

1. Bản chất của sự gắn amino acid vào tRNA

Quá trình gắn amino acid vào tRNA là quá trình hình thành một liên kết acyl giữa nhóm carboxyl của amino acid và nhóm 2'- hoặc 3'-OH của adenine ở đầu 3' của tRNA. Liên kết này được xem là một liên kết giàu năng lượng. Năng lượng giải phóng ra khi liên kết bị phá vỡ giúp hình thành cầu nối peptide để liên kết amino acid với chuỗi polypeptide đang được tổng hợp.

2. Sự nhận diện và gắn amino acid vào tRNA

Sự nhận diện và gắn amino acid vào tRNA tương ứng được thực hiện bởi một enzyme gọi là aminoacyl-tRNA synthetase.

Quá trình này diễn ra như sau: đầu tiên, amino acid được adenylyl hóa bằng cách phản ứng với ATP, kết quả tạo thành amino acid có gắn adenylic acid qua cầu nối ester giàu năng lượng giữa nhóm COOH của amino acid và nhóm phosphoryl của AMP, đồng thời giải phóng ra pyrophosphate. Sau đó, amino acid được adenylyl hóa này (vẫn đang gắn với synthetase) phản ứng tiếp với tRNA. Phản ứng này chuyển amino acid đến đầu 3' của tRNA để gắn với nhóm OH, đồng thời giải phóng AMP.

Phản ứng tổng hợp của quá trình này như sau:

$$\text{Amino acid} + \text{tRNA} + \text{ATP} \rightarrow \text{aminoacyl-tRNA} + \text{AMP} + PP_i$$

3. Tính đặc hiệu của aminoacyl-tRNA synthetase

Hầu hết các tế bào đều có một enzyme synthetase riêng biệt chịu trách nhiệm cho việc gắn một amino acid vào một tRNA tương ứng (như vậy có tất cả 20 synthetase). Tuy nhiên, nhiều vi khuẩn có dưới 20 synthetase. Trong trường hợp này, cùng một synthetase chịu trách nhiệm cho hơn một loại amino acid.

Sự nhận diện amino acid chính xác là dựa vào kích thước, sự tích điện và gốc R khác nhau của các amino acid. Sự nhận diện tRNA dựa vào các trình tự nucleotide khác nhau của tRNA. Tỷ lệ sai sót trong quá trình gắn amino acid với tRNA tương ứng là khá thấp.

4. Phân loại aminoacyl-tRNA synthetase

Có hai loại tRNA synthetase.

- Loại I bao gồm các synthetase gắn các amino acid như Glu, Gln, Arg, Cys, Met, Val, Ile, Leu, Tyr, Trp vào nhóm 2'-OH.
- Loại II gồm các synthetase gắn các amino acid như Gly, Ala, Pro, Ser, Thr, His, Asp, Asn, Lys, Phe vào nhóm 3'-OH.

IV. Các giai đoạn của quá trình dịch mã

Quá trình dịch mã được bắt đầu bằng sự gắn của mRNA và một tRNA khởi đầu với tiểu đơn vị nhỏ tự do của ribosome. Phức hợp tiểu đơn vị nhỏ-mRNA thu hút tiểu đơn vị lớn đến để tạo nên ribosome nguyên vẹn với mRNA được kẹp giữa hai tiểu đơn vị. Sự tổng hợp protein được bắt đầu tại codon khởi đầu ở đầu 5' của mRNA và tiến dần về phía 3'. Khi ribosome dịch mã từ codon này sang codon khác, một tRNA đã gắn amino acid kế tiếp được đưa vào trung tâm giải mã và trung tâm peptidyl transferase của ribosome. Khi ribosome gặp codon kết thúc thì quá trình tổng hợp chuỗi polypeptide kết thúc. Chuỗi này được giải phóng, hai tiểu đơn vị của ribosome rời nhau ra và sẵn sàng đến gặp mRNA mới để thực hiện một chu trình tổng hợp protein mới. Quá trình dịch mã được chia thành ba giai đoạn là khởi đầu, kéo dài và kết thúc.

1. Giai đoạn khởi đầu

1.1. Ở prokaryote

1.1.1. Các yếu tố khởi đầu (IF: initiation factor)

Có các yếu tố khởi đầu xúc tác cho tiểu đơn vị nhỏ trong việc hình thành phức hợp khởi đầu. Đó là IF1, IF2, IF3. Mỗi yếu tố khởi đầu có tác dụng như sau:

- IF1 giúp tiểu đơn vị nhỏ gắn vào mRNA và ngăn cản các tRNA gắn vào vùng thuộc vị trí A trên tiểu đơn vị nhỏ.
- IF2 là một protein gắn và thủy phân GTP. IF2 thúc đẩy sự liên kết giữa fMet-$tRNA_i^{fMet}$ và tiểu đơn vị nhỏ, ngăn cản những aminoacyl-tRNA khác đến gắn vào tiểu đơn vị nhỏ.
- IF3 ngăn cản tiểu đơn vị nhỏ tái liên kết với tiểu đơn vị lớn và gắn với các tRNA mang amino acid. IF3 gắn vào tiểu đơn vị nhỏ vào cuối vòng dịch mã trước, nó giúp tách ribosome 70S thành tiểu đơn vị lớn và tiểu đơn vị nhỏ.

Khi tiểu đơn vị nhỏ đã được gắn ba yếu tố khởi đầu, nó sẽ gắn tRNA khởi đầu và mRNA. Sự gắn hai RNA này là hoàn toàn độc lập với nhau.

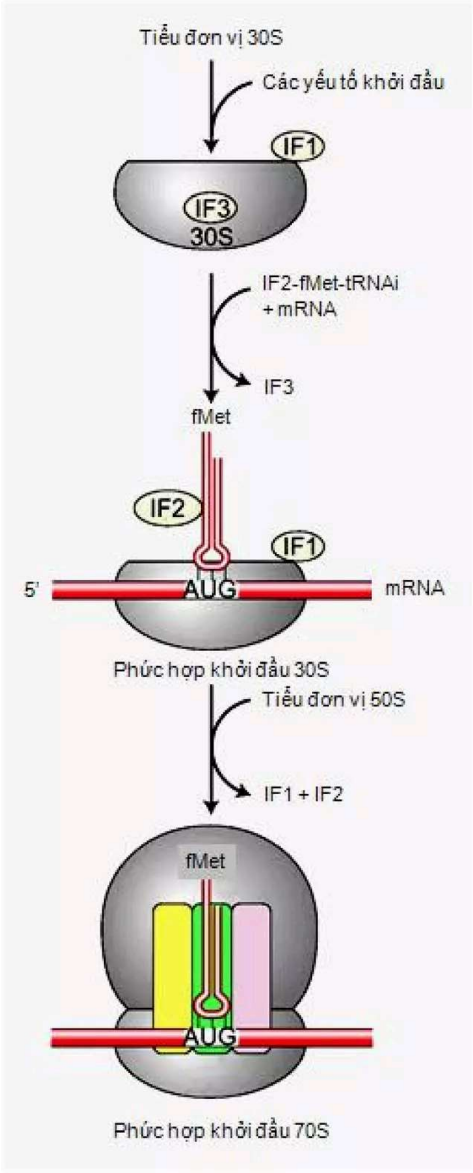
1.1.2. Bước 1: Tiểu đơn vị nhỏ gắn vào codon khởi đầu

Sự liên kết giữa tiểu đơn vị nhỏ với mRNA được thực hiện thông qua sự bắt cặp base bổ sung giữa vị trí gắn ribosome và rRNA 16S. Các mRNA của vi khuẩn có một trình tự nucleotide đặc hiệu gọi là trình tự Shine-Dalgarno (SD) gồm 5-10 nucleotide trước codon khởi đầu. Trình tự này bổ sung với một trình tự nucleotide gần đầu 3' của rRNA 16S. Tiểu đơn vị nhỏ được đặt trên mRNA sao cho codon khởi đầu được đặt đúng vào vị trí P một khi tiểu đơn vị lớn gắn vào phức hợp.

1.1.3. Bước 2: tRNA đầu tiên có mang methionine biến đổi đến gắn trực tiếp với tiểu đơn vị nhỏ

Một tRNA đặc biệt được gọi là tRNA khởi đầu đến gắn trực tiếp với vị trí P (không qua vị trí A). tRNA này có anticodon (bộ ba đối mã) có thể bắt cặp với AUG hoặc GUG. Tuy nhiên tRNA này không mang methionine cũng như valine mà mang một dạng biến đổi của methionine gọi là N-formyl methionine. tRNA khởi đầu này được gọi là fMet-$tRNA_i^{fMet}$.

Trong hoặc sau quá trình tổng hợp polypeptide, gốc formyl được loại bỏ bởi enzyme deformylase. Ngoài ra, aminopeptidase sẽ loại bỏ methionine cũng như một hoặc hai amino acid kế tiếp ở đầu chuỗi polypeptide.



Hình 6.2. Khởi đầu dịch mã ở prokaryote

1.1.4. Bước 3: Hình thành phức hợp khởi đầu 70S

Bước gắn thêm tiểu đơn vị lớn để tạo thành phức hợp khởi đầu 70S diễn ra như sau: khi codon khởi đầu và fMet-$tRNA_i^{fMet}$ bắt cặp với nhau, tiểu đơn vị nhỏ thay đổi hình dạng làm giải phóng IF3. Sự vắng mặt IF3 cho phép tiểu đơn vị lớn gắn vào tiểu đơn vị nhỏ đang mang các thành phần trên. Nhờ có tiểu đơn vị lớn gắn vào, hoạt tính GTPase của IF2-GTP được kích thích để thủy phân GTP. IF2-GDP tạo thành có ái lực thấp đối với ribosome và tRNA khởi đầu dẫn đến sự giải phóng IF2-GDP cũng như IF1. Như vậy phức hợp khởi đầu cuối cùng được tạo thành bao gồm ribosome 70S được gắn tại codon

khởi đầu của mRNA, với fMet-$tRNA_i^{fMet}$ tại vị trí P, còn vị trí A đang trống. Phức hợp này sẵn sàng tiếp nhận một tRNA mang amino acid vào vị trí A để bắt đầu tổng hợp polypeptide (Hình 6.2).

1.2. Ở Eukaryote

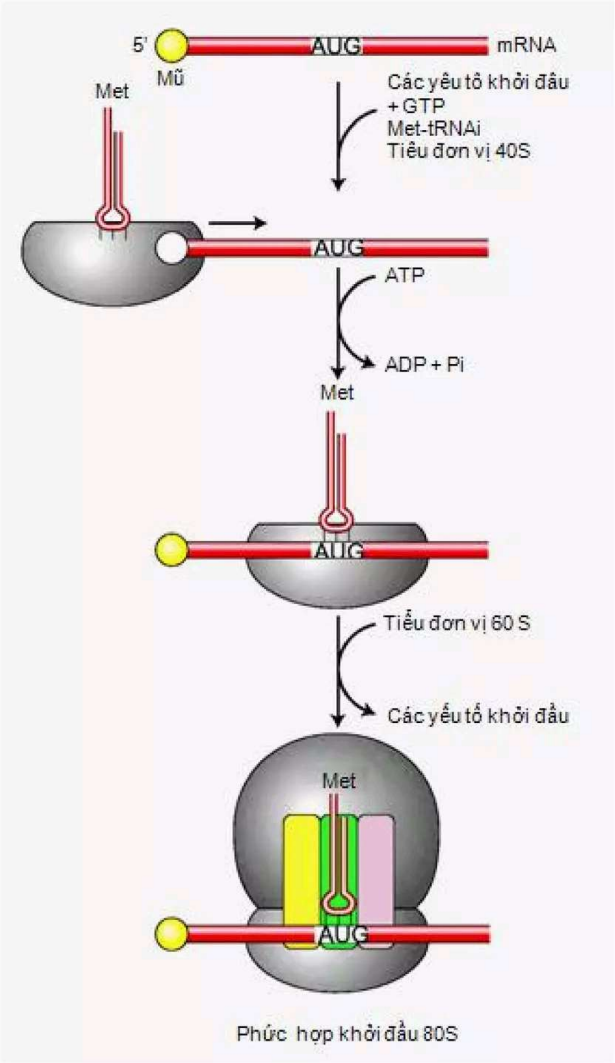
1.2.1. Bước 1: Sự hình thành phức hợp tiền khởi đầu 43S

Giai đoạn khởi đầu đòi hỏi sự hỗ trợ của hơn 30 protein khác nhau, mặc dù eukaryote cũng có những yếu tố khởi đầu tương ứng với prokaryote. Các yếu tố khởi đầu này được ký hiệu là eIF.

Khi ribosome của eukaryote hoàn thành một chu trình dịch mã, nó tách rời ra thành tiểu đơn vị lớn và tiểu đơn vị nhỏ tự do thông qua tác động của các yếu tố eIF3 và eIF1A (tương tự với IF3 ở prokaryote). Hai protein gắn GTP là eIF2 và eIF5B làm trung gian thu hút tRNA khởi đầu đã gắn methionine (chứ không phải N-formyl methionine như ở prokaryote) đến tiểu đơn vị nhỏ. Chính yếu tố eIF5B-GTP là tương đồng với IF2-GTP của prokaryote. Yếu tố này liên kết với tiểu đơn vị nhỏ theo phương thức phụ thuộc eIF1A. Rồi eIF5B-GTP giúp thu hút phức hợp eIF2-GTP và Met-$tRNA_i^{Met}$ đến tiểu đơn vị nhỏ. Hai protein gắn GTP này cùng nhau đưa Met-$tRNA_i^{Met}$ vào vùng thuộc vị trí P của tiểu đơn vị nhỏ. Kết quả, hình thành phức hợp tiền khởi đầu 43S.

1.2.2. Bước 2: Sự nhận dạng mũ 5’ của mRNA

Quá trình này được thực hiện thông qua eIF4F. Yếu tố này có ba tiểu đơn vị, một tiểu đơn vị gắn vào mũ 5’, hai tiểu đơn vị khác gắn với RNA. Phức hợp này lại được gắn với eIF4B làm hoạt hóa một enzyme RNA helicase của một trong những tiểu đơn vị của eIF4F. Helicase này tháo xoắn tất cả các cấu trúc bậc hai được hình thành ở đầu tận cùng của mRNA. Phức hợp eIF4F/B và mRNA lại thu hút phức hợp tiền khởi đầu 43S đến thông qua tương tác giữa eIF4F và eIF3.



Hình 6.3. Khởi đầu dịch mã ở eukaryote

1.2.3. Bước 3: Tiểu đơn vị nhỏ tìm thấy codon khởi đầu bằng cách quét xuôi dòng từ đầu 5' của mRNA và sự hình thành phức hợp khởi đầu 80S

Một khi được gắn vào đầu 5' của mRNA, tiểu đơn vị nhỏ và các yếu tố liên kết với nó di chuyển dọc theo mRNA theo hướng 5' → 3' cho đến khi gặp trình tự 5'-AUG-3' đầu tiên mà nó nhận dạng là codon khởi đầu. Codon được nhận dạng bằng sự bắt cặp base bổ sung giữa anticodon (bộ ba đối mã) của tRNA khởi đầu và codon khởi đầu. Sự bắt cặp này thúc đẩy phóng thích eIF2 và eIF3 cho phép tiểu đơn vị lớn gắn được vào tiểu đơn vị nhỏ. Sự gắn này dẫn đến phóng thích các yếu tố khởi đầu còn lại thông qua sự thủy phân GTP dưới tác dụng của eIF5B. Cuối cùng, Met-tRNA$_i^{Met}$ được đưa vào vị trí P của phức hợp khởi đầu 80S. Lúc này, ribosome ở trong tư thế sẵn sàng tiếp nhận aminoacyl-tRNA vào vị trí A (Hình 6.3).

1.2.4. Những yếu tố khởi đầu dịch mã giữ mRNA eukaryote ở dạng vòng

Ngoài việc gắn vào đầu 5' của mRNA, các yếu tố khởi đầu còn liên kết chặt chẽ với đầu 3' thông qua đuôi poly(A). Điều này được thực hiện bởi sự tương tác giữa eIF4F và protein gắn poly(A) bọc bên ngoài đuôi poly(A). Việc tìm thấy các yếu tố khởi đầu dịch mã có vai trò "vòng hóa" mRNA theo phương thức phụ thuộc đuôi poly(A) đã giải thích được quan sát trước đây là một khi ribosome kết thúc sự dịch mã một mRNA mà được vòng hóa thông qua đuôi poly(A) thì ribosome mới được phóng thích này là ribosome lý tưởng để tái khởi đầu dịch mã trên cùng mRNA.

2. Giai đoạn kéo dài

2.1. Bước 1: Aminoacyl-tRNA được đưa đến vị trí A nhờ yếu tố kéo dài EF-Tu

Một khi tRNA đã gắn amino acid thì EF-Tu đến gắn vào đầu 3' của aminoacyl-tRNA. EF-Tu chỉ có thể gắn với aminoacyl-tRNA khi nó liên kết với GTP. EF-Tu-GTP đưa aminoacyl-tRNA vào vị trí A của ribosome. Chỉ phức hợp aminoacyl-tRNA-EF-Tu-GTP nào có anticodon bổ sung với codon của mRNA tại vị trí A thì mới được giữ lại trên ribosome. Sau đó, EF-Tu tương tác với trung tâm gắn yếu tố của ribosome nằm trên tiểu đơn vị lớn và thủy phân GTP, rồi EF-Tu được phóng thích khỏi tRNA và ribosome, để aminoacyl-tRNA nằm lại tại vị trí A.

2.2. Bước 2: Hình thành cầu nối peptide

Aminoacyl-tRNA tại vị trí A được quay vào trung tâm peptidyl transferase và cầu nối peptide được hình thành. Phản ứng này được xúc tác bởi peptidyl transferase, ngày nay nó được xác định là rRNA, đặc biệt là rRNA 23S của tiểu đơn vị lớn. Vì vậy, peptidyl transferase còn được gọi là ribozyme.

Trong quá trình hình thành cầu nối peptide, cầu nối giữa amino acid và tRNA ở vị trí A không bị phá vỡ. Đầu 3' của cả hai tRNA được đưa đến gần nhau và nhóm amine của amino acid ở vị trí A tấn công nhóm carboxyl của amino acid ở vị trí P. Kết quả là tRNA ở vị trí A mang một dipeptide, trong khi tRNA ở vị trí P đã bị khử acyl.

Sau đó xảy ra sự chuyển dịch (xem bước 3): peptidyl-tRNA (đang mang dipeptide) chuyển sang vị trí P, và vị trí A sẵn sàng tiếp nhận một aminoacyl-tRNA mới. Cầu nối peptide tiếp theo được hình thành theo cách giống hệt trên, trong đó nhóm amine của amino acid mới liên kết với nhóm carboxyl ở đầu C tận cùng của chuỗi polypeptide đang tổng hợp. Thực chất, đây là quá trình chuyển chuỗi polypeptide đang tổng hợp từ peptidyl-tRNA ở vị trí P sang aminoacyl-tRNA ở vị trí A. Vì vậy, phản ứng tạo cầu nối peptide được gọi là phản ứng peptidyl transferase.

Như vậy, chuỗi polypeptide được tổng hợp theo chiều từ đầu N đến đầu C.

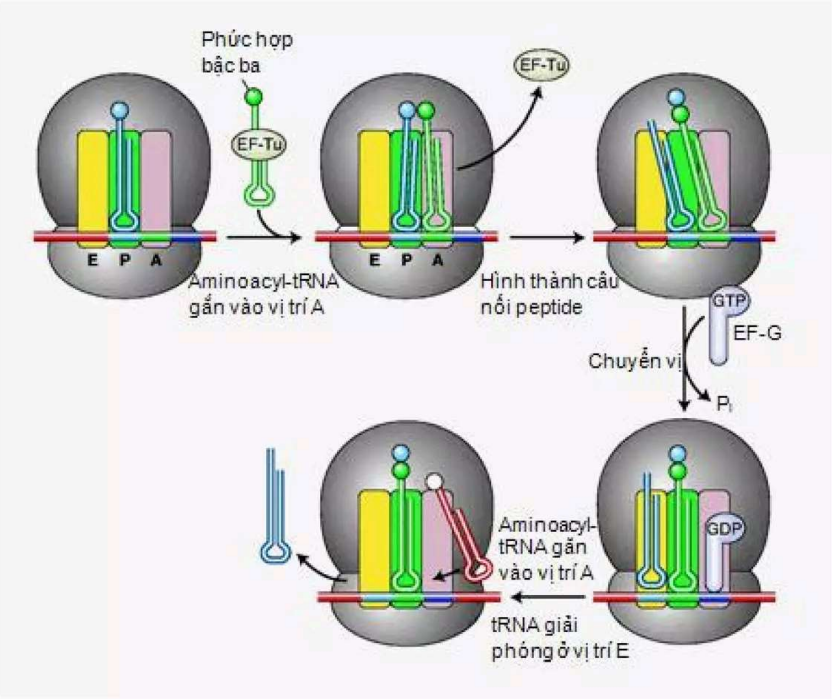
Trong quá trình này, không có sự thủy phân nucleoside triphosphate. Năng lượng được cung cấp từ sự phá vỡ cầu nối acyl giàu năng lượng giữa chuỗi polypeptide đang tổng hợp và tRNA.

2.3. Bước 3: Sự chuyển dịch (translocation)

Một khi phản ứng peptidyl transferase xảy ra thì tRNA trong vị trí P không gắn với amino acid nữa, và chuỗi polypeptide đang hình thành được liên kết với tRNA trong vị trí A. Để một vòng kéo dài polypeptide mới xảy ra, tRNA ở vị trí P phải chuyển đến vị trí E và tRNA ở vị trí A chuyển đến vị trí P. Đồng thời, mRNA phải di chuyển qua 3 nucleotide để ribosome tiếp xúc với codon tiếp theo. Những sự di chuyển này được gọi là sự chuyển dịch.

Bước đầu tiên trong chuyển dịch được song hành với phản ứng peptidyl transferase. Khi chuỗi peptide được chuyển sang tRNA ở vị trí A, đầu 3' của tRNA này hướng đến vùng vị trí P của tiểu đơn vị lớn, trong khi đầu anticodon vẫn còn nằm ở vị trí A. Tương tự, tRNA ở vị trí P (mà không còn gắn chuỗi polypeptide nữa) nằm ở vị trí E của tiểu đơn vị lớn và vị trí P của tiểu đơn vị nhỏ.

Để hoàn thành sự chuyển dịch phải có sự tác động của một yếu tố kéo dài gọi là EF-G. EF-G chỉ gắn vào ribosome khi được liên kết với GTP. Sau khi phản ứng peptidyl transferase xảy ra, sự thay đổi vị trí của tRNA ở vị trí A đã để lộ vị trí gắn cho EF-G. Khi EF-G-GTP gắn vào vị trí này, nó tiếp xúc với trung tâm gắn yếu tố và kích thích thủy phân GTP. Sự thủy phân này làm thay đổi hình dạng của EF-G-GDP và cho phép nó với tới tiểu đơn vị nhỏ để thúc đẩy sự chuyển dịch của tRNA ở vị trí A. Khi sự chuyển dịch được hoàn thành, cấu trúc của ribosome giảm đáng kể ái lực với EF-G-GDP, điều này cho phép yếu tố kéo dài được phóng thích khỏi ribosome. Cùng với việc tRNA ở vị trí A chuyển đến vị trí P, tRNA ở vị trí P chuyển đến vị trí E và mRNA dịch chuyển ba nucleotide. Từ vị trí E, tRNA được phóng thích khỏi ribosome (Hình 6.4).



Hình 6.4. Kéo dài dịch mã

2.4. Các yếu tố kéo dài có gắn GDP (EF-Tu-GDP và EF-G-GDP) được đổi GDP thành GTP trước khi tham gia vào vòng kéo dài mới

EF-Tu và EF-G là những protein xúc tác mà chỉ được sử dụng một lần đối với một vòng kéo dài bao gồm đưa tRNA vào ribosome, hình thành cầu nối peptide, và chuyển dịch. Sau khi GTP được thủy phân, hai protein trên phải giải phóng GDP và gắn với một GTP mới.

Đối với EF-G, do GDP có ái lực thấp với EF-G hơn GTP nên GDP nhanh chóng được giải phóng và GTP mới được gắn vào.

Đối với EF-Tu, cần có sự tham gia của yếu tố hoán đổi GTP gọi là EF-Ts. Sau khi EF-Tu-GDP được phóng thích khỏi ribosome, EF-Ts được gắn vào EF-Tu và thế chỗ của GDP. Sau đó GTP đến gắn vào phức hợp EF-Tu-EF-Ts. Phức hợp sau cùng được tách thành EF-Ts tự do và EF-Tu-GTP.

3. Giai đoạn kết thúc

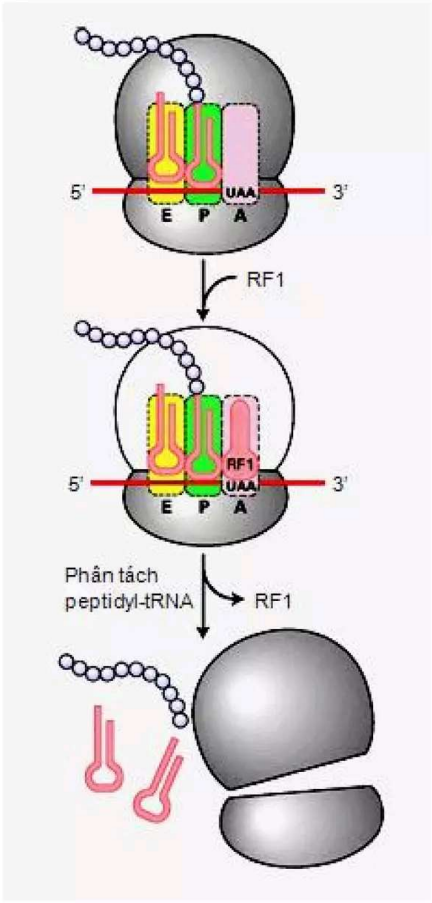
3.1. Các yếu tố giải phóng kết thúc dịch mã

Chu kỳ gắn aminoacyl-tRNA của ribosome, sự hình thành cầu nối peptide, và sự chuyển dịch xảy ra liên tục cho đến khi một trong ba codon kết thúc vào vị trí A. Các codon này được nhận diện bởi các yếu tố giải phóng (RF: release factor) (Hình 6.5).

Có hai loại yếu tố giải phóng:

- Các yếu tố giải phóng loại I nhận diện codon kết thúc và thúc đẩy sự thủy phân để tách chuỗi polypeptide ra khỏi peptidyl-tRNA tại vị trí P. Prokaryote có hai yếu tố giải phóng loại I là RF1 và RF2, trong đó RF1 nhận diện codon kết thúc UAG và RF2 nhận diện UGA, còn UAA được nhận diện bởi cả RF1 và RF2. Eukaryote chỉ có một yếu tố giải phóng gọi là eRF1 nhận diện được cả ba loại codon kết thúc.

- Các yếu tố giải phóng loại II kích thích sự tách yếu tố giải phóng loại I ra khỏi ribosome sau khi chuỗi polypeptide được giải phóng. Chỉ có một yếu tố giải phóng loại II, được gọi là RF3 ở prokaryote và eRF3 ở eukaryote. Yếu tố giải phóng loại II được điều hòa bởi GTP.



Hình 6.5. Kết thúc dịch mã

3.2. Sự hoán đổi GDP/GTP và thủy phân GTP điều khiển hoạt động của yếu tố giải phóng loại II

Yếu tố giải phóng loại II là một protein gắn GTP nhưng có ái lực với GDP cao hơn GTP. Vì vậy, phần lớn RF3 được gắn với GDP. RF3-GDP gắn với ribosome theo một

phương thức phụ thuộc sự hiện diện của yếu tố giải phóng loại I. Sau khi yếu tố giải phóng loại I kích thích sự phóng thích chuỗi polypeptide, xảy ra một sự thay đổi hình dạng ribosome, và yếu tố giải phóng loại I kích thích RF3 hoán đổi GDP thành GTP. Sự gắn GTP vào RF3 dẫn đến sự hình thành tương tác ái lực cao với ribosome và đẩy yếu tố giải phóng loại I ra khỏi ribosome. Sự thay đổi này cho phép RF3 liên kết với trung tâm gắn yếu tố của tiểu đơn vị lớn. Sự tương tác này kích thích thủy phân GTP. Vì không còn yếu tố loại I nữa nên RF3-GDP có ái lực thấp với ribosome và bị phóng thích ra ngoài.

3.3. Sự quay vòng của ribosome

Sau khi phóng thích chuỗi polypeptide và các yếu tố giải phóng, ribosome vẫn còn gắn với mRNA cùng với hai tRNA tại vị trí P và vị trí E. Để ribosome tham gia vào quá trình tổng hợp polypeptide mới, tRNA và mRNA phải đi khỏi ribosome và hai tiểu đơn vị của ribosome phải rời nhau ra. Tập hợp những sự kiện như vậy gọi là sự quay vòng ribosome (ribosome recycling).

Ở prokaryote, có một yếu tố gọi là yếu tố quay vòng ribosome (RRF: ribosome recycling factor). RRF gắn vào vị trí A, nó bắt chước tRNA. RRF lôi kéo EF-G đến ribosome và EF-G kích thích giải phóng những tRNA tại vị trí P và E. Sau đó, EF-G và RRF được phóng thích khỏi ribosome cùng với mRNA. IF3 có thể tham gia vào sự giải phóng mRNA đồng thời nó cũng cần cho sự tách rời hai tiểu đơn vị của ribosome. Kết quả tạo ra tiểu đơn vị nhỏ gắn IF3 và tiểu đơn vị lớn tự do. Ribosome bây giờ có thể tham gia vào vòng dịch mã mới.

V. Các nhân tố ức chế dịch mã

Quá trình dịch mã bao gồm rất nhiều bước và tuân theo một quy tắc là không có một bước nào có thể xảy ra khi bước trước nó chưa hoàn thành. Tuy nhiên, điều này tạo ra một điểm yếu trong dịch mã là nếu chẳng may có một bước không hoàn thành thì toàn bộ quá trình phải bị ngừng lại. Vì vậy, dịch mã thường là điểm đích tác động của nhiều loại kháng sinh và các độc tố (Bảng 6.2).

Khoảng 40% các loại kháng sinh là những yếu tố ức chế dịch mã. Các kháng sinh này tiêu diệt vi khuẩn gây bệnh ở người hoặc động vật do nó tác động làm ngừng dịch mã ở vi khuẩn nhưng không ảnh hưởng đến dịch mã ở eukaryote, do quá trình dịch mã ở vi khuẩn và eukaryote có những điểm khác nhau có ý nghĩa, ví dụ ribosome khác nhau về kích thước và các thành phần. Nhiều kháng sinh gắn chọn lọc với ribosome của vi khuẩn và ức chế nhiều bước trong dịch mã nhưng không ảnh hưởng đến ribosome của eukaryote. Ví dụ tetracycline gắn vào vị trí A của ribosome vi khuẩn và ức chế sự đi vào của các aminoacyl-tRNA nhưng không hiệu quả trên ribosome của eukaryote. Những kháng sinh khác nhau thì ức chế các bước khác nhau trong quá trình dịch mã nên chúng thường được sử dụng để nghiên cứu sự hoạt động của bộ máy dịch mã, đặc biệt là

puromycin. Do cấu trúc ba chiều của puromycin giống với đầu 3' của aminoacyl-tRNA nên nó có thể gắn vào vị trí A và ức chế sự đi vào đây của các aminoacyl-tRNA.

Bảng 6.2. Sự ức chế dịch mã của một số kháng sinh và độc tố

Kháng sinh Độc tố	Tế bào đích	Đích ở mức phân tử	Hậu quả
Tetracycline	Prokaryote	Vị trí A/30S	Ức chế gắn tRNA vào A
Hygromycin B	Prokaryote Eukaryote	Gần vị trí A/30S	Ngăn cản chuyển dịch của tRNA từ A sang P
Paromycin	Prokaryote	Vị trí tương tác codon-anticodon/30S	Giảm tính chọn lọc của sự bắt cặp base giữa codon và anticodon
Chloram-phenicol	Prokaryote	Trung tâm peptidyl transferase/50S	Ngăn cản đưa tRNA tại A vào phản ứng peptidyl transferase chính xác
Puromycin	Prokaryote Eukaryote	Vị trí A/tiểu đơn vị lớn	Kết thúc sớm sự tổng hợp polypeptide
Erythromycin	Prokaryote	Kênh thoát peptide/ 50S	Ức chế chuỗi polypeptide đi ra khỏi ribosome
Fusidic acid	Prokaryote	EF-G	Ngăn cản phóng thích EF-G-GDP khỏi ribosome

Thiostrepton	Prokaryote	Trung tâm gắn yếu tố /50S	Ảnh hưởng sự gắn của IF2 và EF-G với trung tâm này
Kirromycin		EF-Tu	Ngăn cản thay đổi hình dạng liên quan thủy phân GTP, phóng thích EF-Tu
Ricin, α-Sarcin	Prokaryote Eukaryote	Biến đổi hóa học RNA trung tâm gắn yếu tố	Ngăn cản sự hoạt hóa các yếu tố dịch mã GTPase
Diptheria	Eukaryote	Biến đổi hóa học EF-Tu	Ức chế chức năng EF-Tu
Cycloheximide	Eukaryote	Trung tâm peptidyl transferase/60S	Ức chế hoạt tính peptidyl transferase

Cầu nối peptide có thể được hình thành giữa puromycin tại vị trí A và chuỗi polypeptide tại vị trí P. Tuy nhiên, puromycin không gắn vào vị trí P nên sự chuyển dịch không xảy ra và chuỗi polypeptide gắn puromycin rời khỏi ribosome dù quá trình tổng hợp chưa kết thúc. Do cấu trúc của tRNA là tương tự nhau ở mọi sinh vật nên puromycin ức chế dịch mã ở cả vi khuẩn và tế bào eukaryote, hậu quả là nó tiêu diệt cả tế bào vật chủ cùng với vi khuẩn. Thỉnh thoảng, puromycin cũng được dùng để điều trị ung thư nhằm tiêu diệt tế bào khối u.

Tài liệu tham khảo/đọc thêm

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. 2002. Molecular Biology of the Cell. 4th ed. *Garland Science Publishing, Inc.* New York, USA.

2. Griffiths AJF, Wessler SR, Lewontin RC, Gelbart WM, Suzuki DT and Miller JH. 2005. An Introduction to Genetic Analysis. *WH Freeman & Co.* New York, USA.

3. John LB, Carey JC, Bamshad MJ and White RL. 2003. Medical Genetics. *Mosby Publishing,* Missouri, USA.

4. Karp G. 2002. Cell and Molecular Biology. *John Wiley & Sons, Inc.* New York, USA.

5. Pierce BA. 2003. Genetics: A Conceptual Approach. *WH Freeman & Co*. New York, USA.

6. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Losick R. 2004. Molecular Biology of the Gene. *Pearson Education, Inc./Benjamin Cummings Publishing*, San Francisco, USA.

[1]Chính xác là từ 3-4 phân tử RNA và từ 55-82 protein.

Chương 7

Sửa chữa và bảo vệ DNA

Trên phân tử DNA có thể xuất hiện nhiều biến đổi do sai hỏng trong quá trình trao đổi chất, do các tác nhân gây đột biến vật lý và hóa học của môi trường. Tuy nhiên, genome luôn có độ ổn định cao nhờ các cơ chế sửa chữa và bảo vệ DNA. DNA là phân tử duy nhất, mà khi biến đổi hay bị phá hỏng vẫn có khả năng được sửa chữa nhờ tế bào. Các cơ chế sửa sai rất đa dạng và có hiệu quả cao. Ba quá trình bao gồm sửa sai, tái bản và tái tổ hợp DNA liên quan chặt chẽ với nhau. Đây cũng là một minh chứng về sự phối hợp chặt chẽ giữa nhiều cơ chế di truyền.

I. Khái quát về các cơ chế sửa sai

Hầu hết các đột biến trên phân tử DNA thường được khắc phục bằng hai phương thức chính:

- Sửa chữa phục hồi trực tiếp (direct reversal repair), hoặc:

- Cắt bỏ sai hỏng và sửa chữa lại bằng cách dùng trình tự bổ sung (damage excision and repair using complementary sequence).

Sửa chữa trực tiếp thường liên quan đến hai loại sai hỏng trên phân tử DNA do tia tử ngoại gây ra là: cyclobutane-pyrimidine dimer (CPDs) và pyrimidine (6-4) pyrimidone (6-4 PPs). Hai loại sai hỏng này đều làm biến dạng cấu trúc xoắn của DNA. CPDs và 6-4 PPs được nhận biết và sửa chữa nhờ enzyme photolyase. Enzyme này sử dụng năng lượng ánh sáng để thực hiện phản ứng làm thay đổi các liên kết hóa học để nucleotide trở lại dạng bình thường. Phản ứng sửa chữa DNA bằng photolyase xảy ra trong rất nhiều sinh vật prokaryote và eukaryote. Tuy nhiên, quá trình này không thấy ở động vật có vú. Ở người cũng chưa phát hiện được bất kỳ loại photolyase nào.

Đa số sai hỏng của DNA được sửa chữa bằng phương thức thứ hai. Cơ chế này phải sử dụng thông tin di truyền chứa ở một trong hai sợi đơn DNA. Khi trình tự nucleotide trên một sợi bị thay đổi thì sợi thứ hai (liên kết bổ sung với sợi thứ nhất) được dùng làm khuôn mẫu để sửa chữa những sai hỏng đó. Một số cơ chế sửa chữa như sau:

- Hệ thống sửa chữa nhận biết các trình tự DNA không thích hợp với các cặp base chuẩn và thay thế chúng.

- Hệ thống sửa chữa-cắt bỏ (excision-repair system) loại đi một đoạn DNA ở vị trí sai hỏng và sau đó thay thế nó.

- Hệ thống sửa chữa-tái tổ hợp (recombinant-repair system) sử dụng phương thức tái tổ hợp để thay thế vùng sợi đôi bị sai hỏng.

Các hệ thống sửa chữa cũng phức tạp như bộ máy tái bản của nó, điều đó cho thấy tầm quan trọng của chúng đối với sự sống của tế bào. Khi hệ thống sửa chữa phục hồi một sai hỏng của DNA, thì không có một hậu quả xấu nào xảy ra. Nhưng một đột biến có thể tạo ra hậu quả xấu khi DNA bị hỏng.

Hình 7.1 tóm tắt một số cơ chế sửa chữa DNA như sau:

- Một vài enzyme phục hồi trực tiếp các loại sai hỏng đặc biệt của DNA.

- Một số phương thức sửa chữa bằng cách cắt bỏ base, sửa chữa bằng cách cắt bỏ nucleotide, và sửa chữa ghép đôi lệch, tất cả chức năng của chúng thực hiện bằng cách loại bỏ và thay thế nguyên liệu.

- Các hệ thống chức năng có thể tái tổ hợp để tạo ra một bản sao không bị sai hỏng thay thế cho một sợi đôi bị sai hỏng.

- Phương thức nối đầu không tương đồng đã nối lại các đầu sợi đôi bị hỏng.

- Một số DNA polymerase khác nhau cần thiết trong việc tổng hợp lại các đoạn DNA thay thế.

1. Các biến đổi xảy ra trên phân tử DNA

Trên DNA có thể xảy ra các biến đổi ngẫu nhiên như sau:

- Gãy hay đứt mạch. Phân tử DNA có chiều ngang rất mảnh, bản thân nó lại thường xuyên cuộn xoắn và giãn xoắn nên dễ xảy ra đứt gãy một sợi. Tuy nhiên, khả năng đứt cùng lúc hai sợi hiếm khi gặp hơn.

Hình 7.1. Việc sửa chữa các gen có thể được phân loại theo các phương thức sử dụng các cơ chế khác nhau để phục hồi hoặc bỏ qua sự sai hỏng của DNA

- Base bị cắt mất. Hiện tượng này làm base tương ứng không bắt cặp được. Dưới tác dụng của nhiệt có thể xảy ra quá trình khử purine (depurination) do thủy phân liên kết *N*-glycosyl.

- Gắn các nhóm mới vào base bằng liên kết cộng hóa trị làm thay đổi tính chất như trường hợp methyl hóa (gắn nhóm CH_3 vào một base).

- Biến một base này thành một base khác làm bắt cặp sai. Ví dụ: quá trình khử amine (desamination) của cytosine biến nó thành uracil.

- Các base có thể tồn tại ở hai dạng keto và enol nên có thể dẫn đến bắt cặp sai. Ví dụ: dạng enol của cytosine có thể bắt cặp với adenine.

- Tạo các thymine dimer.

- Liên kết chéo giữa các mạch (interstand crosslink).

Trên đây là các biến đổi ngẫu nhiên, nếu có tác động của các tác nhân gây đột biến các biến đổi sẽ xảy ra nhiều hơn.

2. Khái quát các cơ chế sửa chữa ở mức phân tử

Sự nguyên vẹn của phân tử DNA mang thông tin di truyền có ý nghĩa sống còn đối với tế bào, nhất là tế bào prokaryote. Ở tế bào vi khuẩn, có đến 50% DNA không bị biến đổi thậm chí sau khi tái bản đến 100 triệu lần. Sự ổn định cao của DNA tế bào có được nhờ hàng loạt các cơ chế bảo vệ sự nguyên vẹn và sửa chữa ngay lập tức bất kỳ sai hỏng nào vừa xuất hiện.

Các hệ thống sửa sai rất đa dạng và có hiệu quả rất cao ở *E. coli*. Có khoảng 100 locus tham gia trực tiếp hoặc gián tiếp vào việc bảo vệ DNA và sửa sai.

Chỉ với DNA, tế bào phải đầu tư rất lớn cho sự ổn định của thông tin di truyền. Nếu như sự tái bản chỉ xảy ra khi phân bào, thì các cơ chế sửa sai phải hoạt động liên tục. Tái bản được thực hiện với cơ chế về cơ bản giống nhau ở các loại tế bào và trong mỗi lần nhân đôi DNA. Trong khi đó, sai hỏng xảy ra rất đa dạng, nên các cơ chế sửa sai nhiều hơn và với số lượng gen tham gia lớn hơn. Hơn nữa, các cơ chế di truyền cơ bản như tái bản, tái tổ hợp DNA đều có sự tham gia của các cơ chế sửa sai và ngược lại sửa sai cần đến tái bản và tái tổ hợp.

3. Biến đổi làm tăng tần số đột biến

Nhờ các cơ chế sửa sai nên bình thường DNA tế bào có độ ổn định rất lớn, thường sự thay thế base có tần số 10^{-9}-10^{-10}. Tuy nhiên, người ta đã phát hiện được các dòng đột biến ngẫu nhiên tăng cao hơn hẳn, chúng được gọi là mutator (nhân tố gây đột biến). Trong nhiều trường hợp, các kiểu hình mutator liên quan đến sai hỏng trong hệ thống sửa sai. Ở *E. coli*, các locus mutator *mutH*, *mutL*, *mutU* và *mutS* tác động đến các cấu phần của hệ thống sửa sai do ghép đôi lệch (mismatch-repair) sau tái bản nên đã làm cho tần số đột biến ngẫu nhiên tăng cao.

Ngoài mutator, sự sai hỏng làm tăng tính nhạy cảm với tia tử ngoại (ultraviolet) và tăng sai hỏng khi tái tổ hợp.

II. Các kiểu sửa chữa

1. Quang tái hoạt hóa

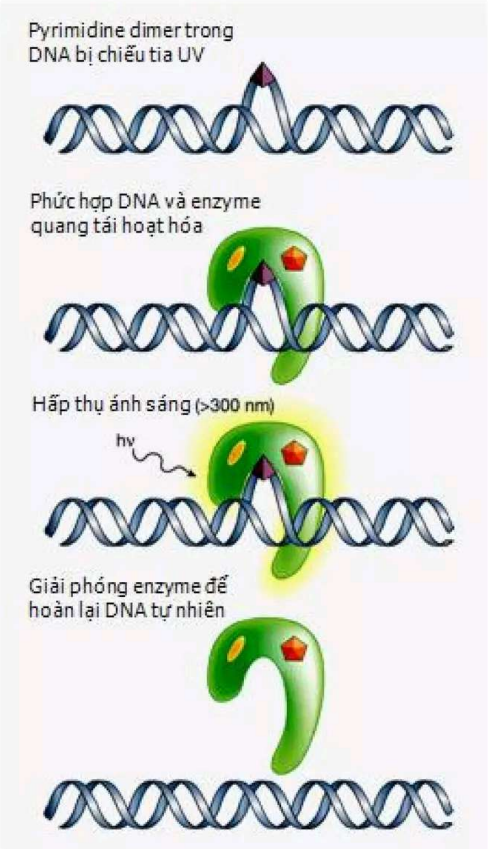
Sửa chữa trực tiếp bằng quang tái hoạt hóa (photoreactivation) ít khi gặp, nó bao gồm sự phục hồi hoặc loại bỏ đơn giản các sai hỏng và quá trình này xảy ra ở ngoài sáng. Quang tái hoạt hóa của các pyrimidine dimers, trong đó tạo cơ hội cho các liên kết cộng hóa trị được phục hồi nhờ enzyme phụ thuộc ánh sáng (light-dependent enzyme), là một ví dụ điển hình (Hình 7.2). Hệ thống sửa chữa này phổ biến trong tự nhiên, và đặc biệt quan trọng ở thực vật. Trong *E. coli*, nó phụ thuộc vào sản phẩm của một gen đơn (*phr*) mã hóa cho một enzyme được gọi là photolyase.

Sau khi xử lý tia tử ngoại gây đột biến, nếu đưa ra ánh sáng thì phần lớn sai hỏng được phục hồi. Hiện tượng này được gọi là quang tái hoạt hóa. Năng lượng của ánh sáng khả kiến (từ 300 đến 600 nm) hoạt hóa photolyase (gen *phr* ở *E. coli*) cắt các vòng cyclobutyl pyrimidine dimer (thường là thymine dimer).

Enzyme này hoạt động trong các tế bào ở nhiều loài khác nhau. Vào ban ngày, các sinh vật thường chịu tác động của ánh sáng, nên cơ chế quang tái hoạt hóa (quang phục hồi) có vai trò quan trọng trong sửa sai DNA. Ví dụ: *Mycoplasma*, sinh vật đơn giản nhất hiện nay, chỉ có vài trăm gen nhưng một trong số đó đã được dùng cho quang tái hoạt hóa.

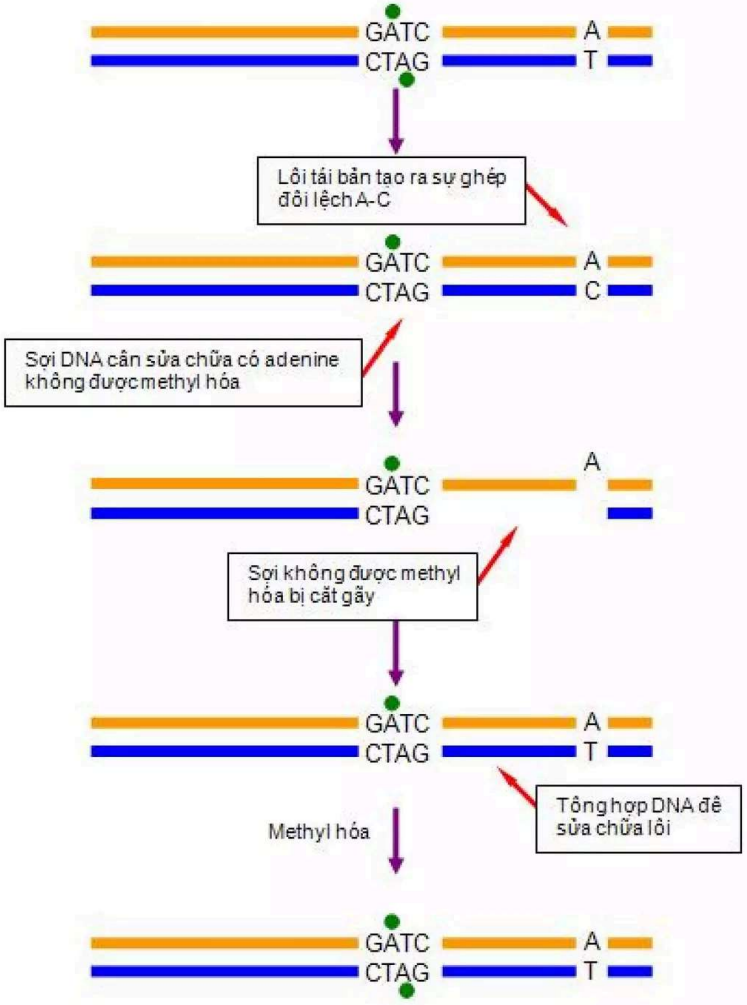
2. Sửa chữa ghép đôi lệch

Sửa chữa ghép đôi lệch (mismatch-repair) giữa các sợi DNA là một trong những mục tiêu chính của hệ thống sửa sai. Sửa chữa ghép đôi lệch được tiến hành khi phát hiện trên DNA các base nằm cạnh nhau mà không bắt cặp thích hợp. Các ghép đôi lệch tăng lên trong suốt quá trình tái bản được sửa chữa bằng cách phân biệt giữa các sợi cũ và mới và được ưu tiên sửa chữa trình tự của các sợi được tổng hợp mới (Hình 7.3). Ghép đôi lệch được tạo ra bởi các biến đổi base, là một kết quả của sự khử amine (deamination).



Hình 7.2. Quang tái hoạt hóa

Ghép đôi lệch thường được sửa sai bằng phương thức sửa chữa cắt bỏ, được khởi đầu bằng một enzyme nhận biết một base sai hỏng thực sự hoặc có sự thay đổi chiều hướng không gian (spatial path) của DNA.



Hình 7.3. Sửa chữa ghép đôi lệch trong tái bản

Sự nhận biết và sửa chữa các sai lệch do bắt cặp base sai trên DNA được phát hiện ở *E. coli*, nấm men và tế bào động vật có vú. Ở vi khuẩn *E. coli*, có ba hệ thống enzyme khác nhau được sử dụng để sửa chữa các sai lệch bằng cách:

- Loại bỏ chỗ sai trong tái bản (errors in replication).

- Loại bỏ ghép đôi lệch bên trong các đoạn trung gian của tái tổ hợp (mismatch within recombinant intermediates).

- Loại bỏ thymine của cặp GT trong trường hợp 5-methylcytosine bị mất nhóm amine (NH_2) thành thymine.

3. Sửa chữa cắt bỏ

Có hai loại hệ thống sửa chữa cắt bỏ, đó là:

- Hệ thống sửa chữa cắt bỏ base (base excision repair) loại bỏ trực tiếp base sai hỏng và thay thế nó trong DNA. Ví dụ điển hình là enzyme DNA uracil glycolase, loại các uracil không được bắt cặp (mispaired) với các guanine.

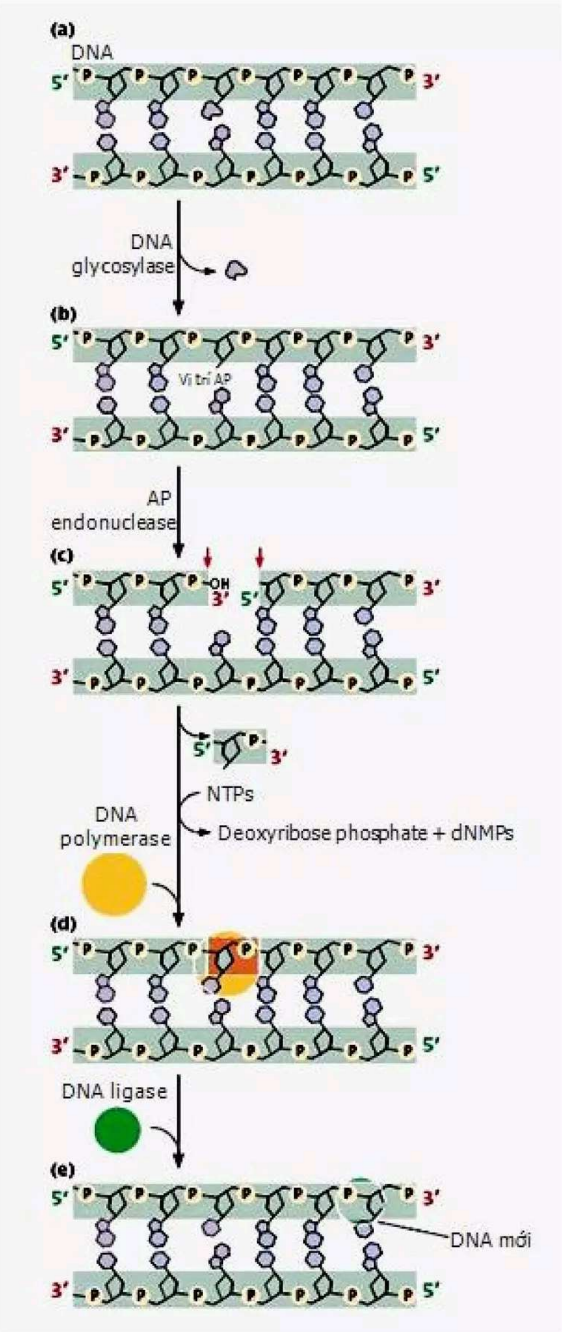
- Hệ thống sửa chữa cắt bỏ nucleotide (nucleotide excision repair) cắt bỏ một trình tự bao gồm các base sai hỏng; sau đó một đoạn DNA mới được tổng hợp để thay thế các nguyên liệu bị cắt bỏ. Các hệ thống này phổ biến ở hầu hết sinh vật.

3.1. Cắt bỏ base

Việc loại bỏ chỗ sai được thực hiện nhờ một hoặc vài enzyme *N*-glycosylase. *N*-glycosylase nhận biết base biến đổi hay mất gốc amine hoặc biến dạng cấu trúc xoắn do sai lệch tạo ra và thủy phân liên kết *N*-glycosylic giữa base với đường với pentose. Lỗ hổng vừa được tạo ra do cắt bỏ được DNA polymerase làm đầy lại dựa vào khuôn mẫu bổ sung đối diện và ligase nối liền chúng (Hình 7.4).

3.2. Cắt bỏ nucleotide

Việc cắt bỏ một vùng có nhiều thymine dimer (do UV) hoặc vùng liên kết chéo giữa các sợi, được thực hiện nhờ incision nuclease (nuclease tạo khấc trên DNA) như phức hợp Uvr ABC của *E. coli,* phức hợp này cắt các đoạn 12-13 nucleotide từ một sợi DNA. Sự thủy phân được thực hiện ở liên kết phosphodiester thứ tám ở đầu 5’ đến chỗ hỏng và ở phía 3’ là liên kết thứ tư hay năm.



Hình 7.4. Phương thức cắt bỏ base

Cấu trúc sợi kép bổ sung cho nhau của DNA cho phép, nếu thông tin bị mất do cắt bỏ sai hỏng trên một sợi có thể chép lại được nhờ dựa vào sợi đơn bổ sung kia. Tuy nhiên, một số sai hỏng liên quan cùng lúc cả hai sợi cũng có thể được sửa chữa nhờ vào

một đoạn nucleotide tương đồng tồn tại đâu đó trong genome. Sự mất đoạn hay xen đoạn nucleotide, hay liên kết chéo giữa các sợi DNA có thể được phục hồi nhờ sự thay thế đoạn tương ứng bằng tái tổ hợp. Thậm chí sự đứt đôi cả hai sợi cùng chỗ, được coi là nặng nhất trong các sai hỏng, có thể được gắn liền lại nhờ ligase hay tái tổ hợp (Hình 7.5).

4. Đọc sửa đối với các base bắt cặp sai

Cơ chế đọc sửa đối với các base bắt cặp sai (proofreading for base-pair matching) được thực hiện trong tái bản DNA. Sự bắt cặp cẩn thận của DNA polymerase đảm bảo cho sự chính xác của tái bản. Trước khi thực hiện phản ứng polymer hóa nối với nucleotide, nucleotide triphosphate mới phải bắt cặp với base bổ sung trên sợi khuôn mẫu. Nếu sự bắt cặp sai xảy ra, DNA polymerase sẽ loại bỏ nucleotide bắt cặp sai.

Thậm chí, trước khi nucleotide mới gắn vào, enzyme dò lại cặp base cuối nếu chúng không bắt cặp thì sự polymer hóa tiếp theo bị dừng. Cặp nucleotide ở đầu cuối 3' bắt cặp sai sẽ bị loại bỏ bởi hoạt tính exonuclease 3'→5' của DNA polymerase. Khi sự bắt cặp của cấu trúc sợi kép đã đúng, quá trình polymer hóa mới được tiếp tục.

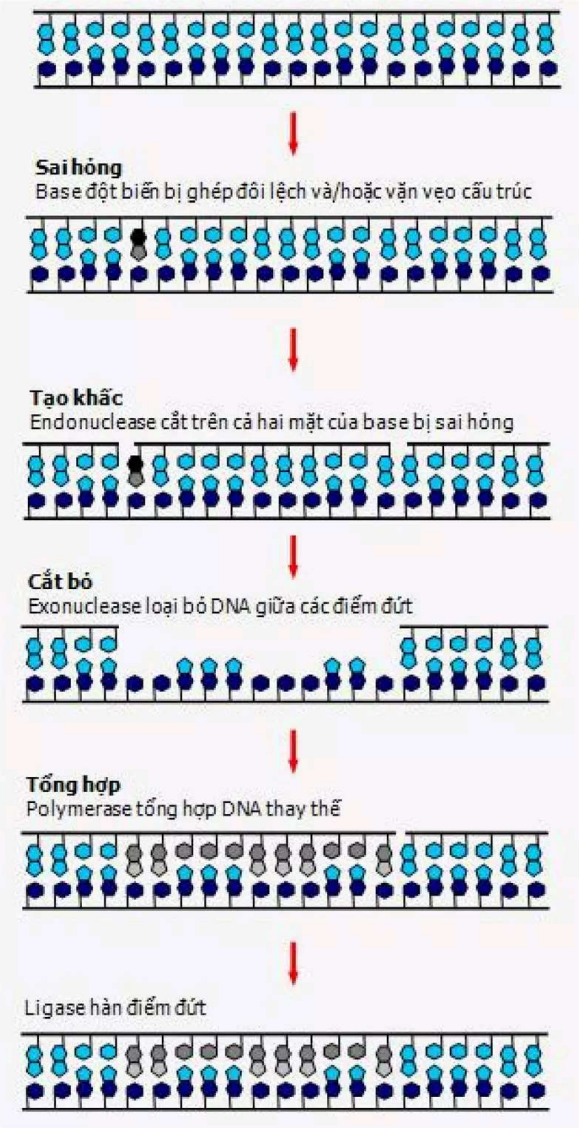
Hoạt tính đọc sửa đối với các base bắt cặp sai là đặc tính của nhiều DNA polymerase đảm bảo cho sự kéo dài chính xác của sợi đang được tổng hợp. Tuy nhiên, trong trường hợp cần thiết, DNA polymerase có thể bỏ qua chỗ sai và chép tiếp nhờ cơ chế tái bản "error prone" (xem mục 6).

5. Các hệ thống sửa chữa tái tổ hợp

Các vị trí có sự sai hỏng, xảy ra trong một phân tử con (daughter molecule), sẽ được phục hồi nhờ sự tái tổ hợp để thu được một bản sao khác của trình tự từ một nguồn không bị sai hỏng. Bản sao này sau đó được dùng để sửa chữa lỗ hổng (gap) trên sợi bị hỏng.

Khi DNA polymerase gặp thymine dimer hay một số sai lệch khác trên sợi DNA khuôn mẫu, có hai khả năng xảy ra:

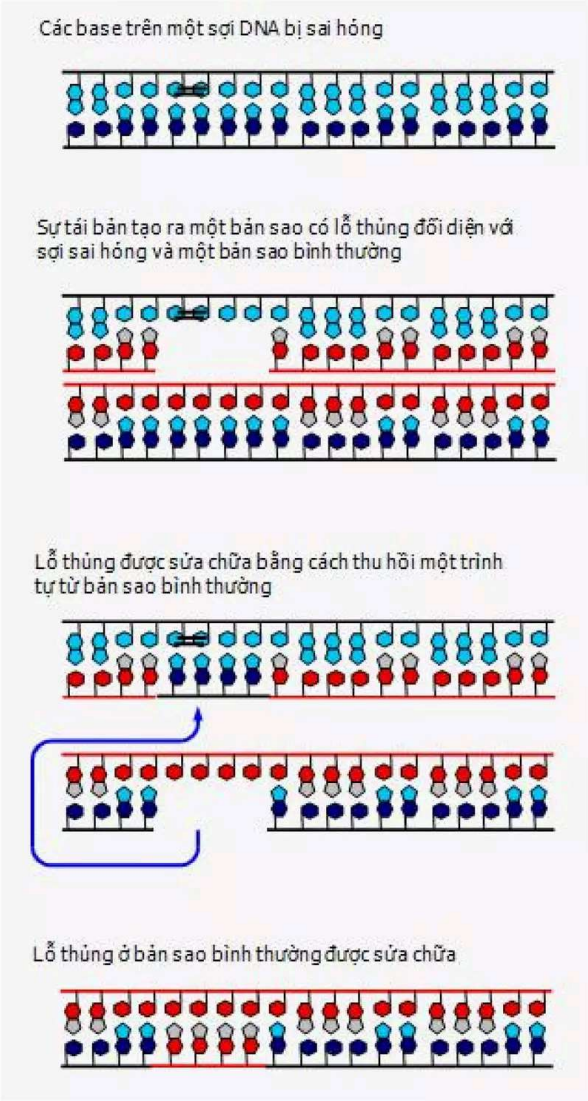
- Polymerase lấp dần lỗ hổng bằng tái bản không theo khuôn mẫu do tái tổ hợp và vượt qua chỗ sai.



Hình 7.5. Cắt bỏ nucleotide. Sửa chữa và cắt bỏ một đoạn DNA có (các) base sai hỏng.

- Polymerase dừng lại và huy động khoảng 1.000 nucleotide phía dưới (downstream); chỗ trống gián đoạn được lấp đầy với sợi bổ sung từ sợi đôi chị em (sister duplex) do tái tổ hợp.

Trong cả hai trường hợp, chỗ sai lệch vẫn còn và được cắt bỏ sau đó bởi sửa sai trực tiếp hay cắt bỏ.



Hình 7.6. Sửa sai nhờ tái tổ hợp và tái bản

6. Hệ thống SOS

Hệ thống SOS (chứa khoảng 30 gen không liên kết và thường bị ức chế bởi protein LexA) sẽ hoạt động khi tế bào bị các tác nhân gây đột biến tạo nhiều sai hỏng trên DNA. Trong trường hợp DNA bị hỏng ngừng tái bản, phản ứng SOS sẽ phục hồi tái bản bằng phương thức tái bản "error-prone".

Ở vi khuẩn *E. coli* người ta đã quan sát thấy sự phá hủy DNA làm mở ra khoảng 20 gen của hệ thống SOS, được kiểm soát âm bởi chất ức chế LexA (LexA repressor).

Trong trường hợp có nhiều sai hỏng cần cấp cứu, LexA bị kích thích, thay đổi cấu hình, tự cắt và mất hoạt tính ức chế. Lúc đó các gen của hệ thống SOS được mở ra. Nếu quá trình sửa sai thực hiện không kịp thì tế bào phải chấp nhận hoặc bị đột biến hoặc chết.

Protein RecA khởi động hệ thống SOS như sau:

- Hộp SOS là trình tự DNA (operator) dài khoảng 20 bp được nhận biết bởi protein của chất ức chế LexA.

- Sai hỏng của DNA làm cho RecA khởi động phản ứng SOS, chứa các gen mã hóa cho nhiều enzyme sửa chữa.

- RecA hoạt hóa hoạt tính tự phân giải tự của LexA.

- LexA ngăn chặn hệ thống SOS, nhưng phản ứng tự phân giải của nó đã hoạt hóa các gen của hệ thống này.

6.1. Sửa chữa theo cơ chế "error-prone"

Phân tử DNA có thể bị sai hỏng nặng nề khi bị chiếu tia tử ngoại hoặc khi chịu tác động của các tác nhân gây ung thư (carcinogens). Lúc đó, không phải chỉ một sợi đơn DNA bị hỏng mà cả hai sợi đều bị đứt gãy và mất đi một số nucleotide. Thông tin di truyền của đoạn hỏng bị mất hoàn toàn nên không có sợi khuôn mẫu để sửa chữa theo các cơ chế thông thường. Trong trường hợp này, tế bào còn lại một giải pháp duy nhất để tăng khả năng sống sót đó là sửa chữa DNA một cách ngẫu nhiên với tỷ lệ sai sót (đột biến) cao nhưng dù sao vẫn duy trì được sự sống. Đây chính là cơ chế sửa chữa DNA "error-prone" (còn được xem như là sự phát sinh đột biến SOS). Như vậy, sửa chữa DNA theo cơ chế này cũng là một trong những nguyên nhân tạo nên tính đa dạng của DNA.

Khi phân tử DNA bị đột biến, một số protein đặc biệt sẽ làm nhiệm vụ dừng quá trình tổng hợp DNA để sửa chữa theo các cơ chế khác nhau. Bởi vì nếu tế bào tiếp tục phân chia (tức là DNA tiếp tục được nhân lên) thì tỷ lệ đột biến cao sẽ xuất hiện khiến số tế bào sống sót giảm xuống.

Một số trường hợp sai hỏng được sửa chữa bằng cơ chế error-prone:

- DNA kết hợp với các base không bổ sung trên sợi con (bắt cặp sai).

- DNA bị sai hỏng không được sửa chữa do DNA polymerase III bị giữ lại trong suốt quá trình tái bản.

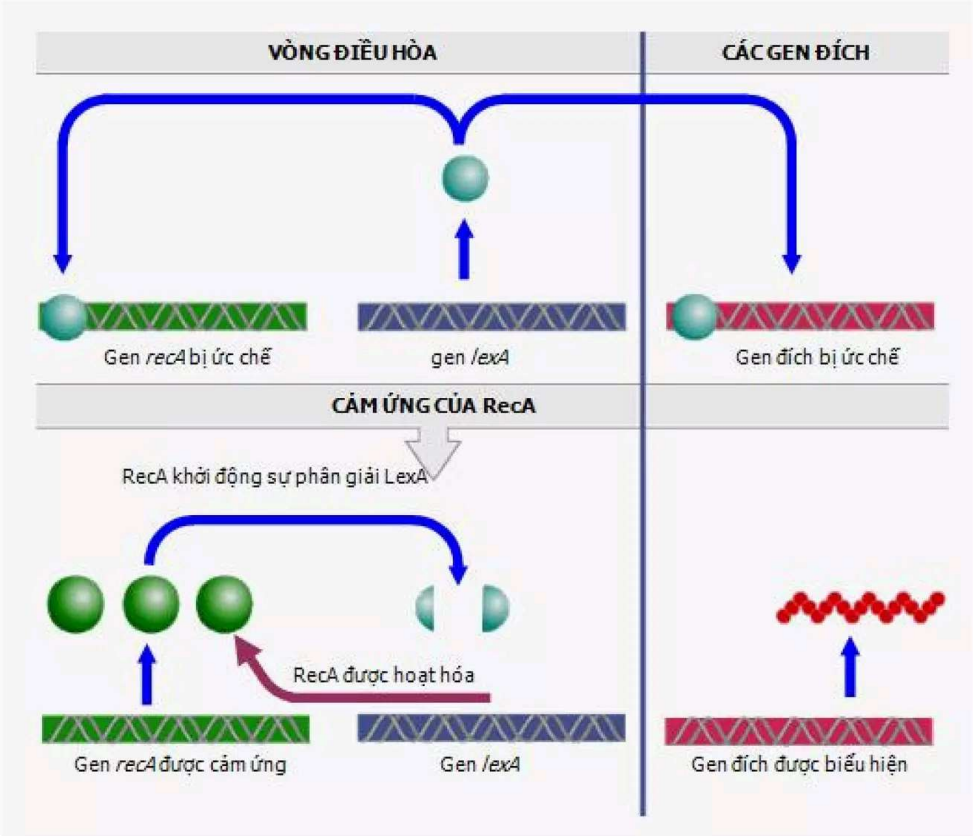
DNA polymerase V (được mã hóa bởi gen *umuD* và *umuC* của *E. coli*), hoặc DNA polymerase IV (được mã hóa bởi gen *dinB* của *E. coli*) được cảm ứng trong hệ thống SOS, để tổng hợp một đoạn bổ sung cho sợi DNA bị hỏng. DNA polymerase V là một phức hợp chứa hai protein tiểu đơn vị UmuD' và một protein tiểu đơn vị UmuC (Trong quá trình sửa chữa "error-prone", protein RecA, hoạt động như một co-protease, gián tiếp tự xúc tác cắt protein UmuD thành một đoạn có hoạt tính, UmuD'). DNA polymerase IV

và V là một phần của một họ lớn, bao gồm các DNA polymerase là những enzyme cần thiết trong sửa chữa DNA bị sai hỏng do chúng có hoạt tính polymerase.

Các protein tiểu đơn vị UmuD (thực tế là UmuD') và UmuC có khả năng gắn các nucleotide vào sợi DNA bất chấp không tồn tại sợi khuôn mẫu. Như vậy, chúng có thể ức chế hoạt tính tự sửa exonuclease 3'→5' của DNA polymerase hoặc chính chúng là một loại DNA polymerase đặc biệt (DNA polymerase V). Khi đột biến xảy ra trên gen mã hóa cho UmuD và UmuC, các tế bào vi khuẩn sẽ sống sót ít hơn tế bào bình thường dưới tác dụng của tia tử ngoại. Một vài plasmid mang gen *mucA* và *mucB*, tương đồng với gen *umuD* và gen *umuC*, khi được đưa vào trong vi khuẩn sẽ làm tăng tính kháng với tia tử ngoại. Ngoài ra, khi cả hai sợi đơn DNA đều chịu đột biến nhưng nếu trong genome còn có ít nhất một bản sao giống hệt đoạn bị hỏng, lúc đó đoạn bị hỏng có thể được phục hồi nhờ cơ chế trao đổi chéo dùng bản sao làm khuôn mẫu.

6.2. Protein LexA

Hoạt tính của RecA gây ra sự phân giải sản phẩm protein của gen *lexA*. LexA là một protein nhỏ (22 kDa) khá ổn định trong các tế bào không bị xử lý, nơi nó có chức năng như một chất ức chế ở nhiều operon. Phản ứng phân giải là không thường xuyên; LexA có hoạt tính protease muộn được hoạt hóa bởi RecA. Khi RecA được hoạt hóa, nó sẽ gây ra phản ứng tự phân giải của LexA làm bất hoạt chức năng ức chế của LexA, và cảm ứng phối hợp tất cả operon mà nó được liên kết. Phương thức này được minh họa ở hình 7.7.



Hình 7.7. LexA và RecA có quan hệ đối kháng thuận nghịch. Protein LexA ức chế nhiều gen có các chức năng sửa chữa, *recA* và *lexA*. Hoạt tính của RecA dẫn đến phân giải protein LexA và cảm ứng tất cả các gen này.

Các gen đích của protein ức chế LexA có nhiều chức năng sửa chữa. Một số gen SOS này chỉ hoạt động ở các tế bào bị xử lý. Những gen khác hoạt động ở các tế bào không bị xử lý, nhưng mức độ biểu hiện được tăng lên nhờ sự phân giải của LexA. Sau khi phân giải LexA, gen có thể được biểu hiện từ promoter thứ hai giống như promoter đầu tiên.

Protein LexA ức chế các gen đích của nó bằng cách liên kết với một đoạn DNA dài 20 bp được gọi là hộp SOS, bao gồm một trình tự liên ứng (consensus sequence: 5'-CTGTN$_8$ACAG-3') với 8 vị trí bảo toàn tuyệt đối (N$_8$). Giống như các operator khác, các hộp SOS xếp chồng lên các promoter tương ứng. Ở locus *lexA*, đối tượng của sự ức chế tự sinh, có hai hộp SOS ở cạnh nhau.

6.3. Protein RecA

Sự cuộn lại trực tiếp của protein RecA trong sửa chữa-tái tổ hợp chỉ là một trong những hoạt tính của nó. Nó có thể được hoạt hóa bởi các xử lý làm sai hỏng DNA hoặc ức chế sự tái bản trong *E. coli*. Điều này giúp nó khởi động một chuỗi phức tạp những thay đổi kiểu hình được gọi là phản ứng SOS, yếu tố cần thiết cho sự biểu hiện của nhiều gen mà các sản phẩm của chúng có các chức năng sửa chữa. Các hoạt tính kép này của protein RecA làm cho người ta không biết rõ sự thiếu hụt trong sửa chữa ở các tế bào đột biến *recA* là do mất chức năng trao đổi sợi DNA của RecA hay là do một vài chức năng khác mà sự cảm ứng của nó phụ thuộc vào hoạt tính protease.

Sự sai hỏng xuất hiện có thể do chiếu tia UV (hầu hết trong các trường hợp nghiên cứu) hoặc do các nhân tố liên kết chéo hoặc sự alkyl hóa (alkylation). Sự ức chế tái bản do một số phương thức như thiếu hụt (deprivation) thymine, bổ sung thêm thuốc, hoặc các đột biến trong một số gen *dna*, có cùng hiệu quả.

Đầu tiên là hoạt hóa RecA bằng xử lý sai hỏng. Chúng ta không biết nhiều về mối quan hệ giữa sự sai hỏng và sự thay đổi đột ngột trong hoạt tính RecA. Một biến động của các trường hợp sai hỏng có thể cảm ứng phản ứng SOS. Hiện nay, người ta thiên về ý kiến cho rằng RecA được hoạt hóa bởi một số chất trung gian trong chuyển hóa DNA.

Tín hiệu cảm ứng có thể chứa một phân tử nhỏ được phóng thích từ DNA; hoặc nó có thể là một vài cấu trúc được tạo thành trong chính DNA. Ở điều kiện *in vitro*, hoạt tính của RecA đòi hỏi sự có mặt của DNA sợi đơn và ATP. Vì thế, một vài tín hiệu có thể hiện diện ở vùng sợi đơn ở một vị trí sai hỏng. Sự tương tác của nó với RecA là nhanh: phản ứng SOS xảy ra trong một vài phút xử lý sai hỏng.

Tóm lại. RecA và LexA là các mục tiêu chung trong hộp SOS: RecA khởi động sự phân giải của LexA, nhân tố ức chế *recA* và chính nó. Vì thế, phản ứng SOS tạo ra sự khuếch đại của cả hai protein RecA và chất ức chế LexA. Kết quả là không còn mâu thuẫn như lúc đầu nữa.

Việc tăng mức độ biểu hiện của protein RecA cần thiết cho vai trò trực tiếp của nó trong các phương thức sửa chữa-tái tổ hợp. Về sự cảm ứng, nồng độ của RecA được tăng lên 50 lần so với nồng độ ban đầu của nó khoảng 1.200 phân tử/tế bào. Nồng độ cao ở các tế bào được cảm ứng có nghĩa là có đủ RecA để đảm bảo rằng tất cả protein LexA bị phân giải.

III. Bảo vệ DNA: Hệ thống cắt hạn chế (R)-Biến đổi (M)

Ngoài các hệ thống sửa sai, tế bào còn có hệ thống bảo vệ DNA. Các vi khuẩn có hệ thống miễn nhiễm hiệu quả DNA lạ, đó là: hệ thống cắt hạn chế-biến đổi (restriction-modification system) hiện diện trong nhiều loài vi khuẩn, trong đó DNA được biến đổi bởi các enzyme đặc hiệu.

Sự biến đổi xảy ra nhằm mục đích bảo vệ DNA chống lại sự phân hủy enzyme (cắt hạn chế) bởi các endonuclease của chính chúng. Sự biến đổi bao gồm một kiểu đặc trưng của phản ứng methyl hóa các gốc nucleotide trong DNA. Bằng cách dùng *S*-adenosylmethionine như là một chất cho, các enzyme methyl hóa (methylase) sẽ hoạt động trên DNA sợi đôi (dsDNA). Các vị trí mà ở đó có sự biến đổi cũng được nhận biết bởi các enzyme hạn chế (restriction enzyme, RE) tương ứng, tuy nhiên các enzyme hạn chế không thể cắt DNA trong các vị trí đã được biến đổi này ở một hoặc cả hai sợi DNA.

Các vi sinh vật prokaryote lẫn eukaryote đều có các enzyme methyl hóa gắn nhóm CH_3 ở những điểm nhất định trên phân tử DNA. Các enzyme này có tính đặc hiệu cao chuyên hóa cho từng dòng vi khuẩn, vì thế DNA của mỗi dòng vi khuẩn chỉ được methyl hóa ở những vị trí đặc biệt nhất định. Nhờ đó, mỗi vi khuẩn dễ dàng phân biệt DNA của bản thân chúng với DNA ngoại lai (foreign DNA) (ví dụ: DNA của bacteriophage) xâm nhập vào trong tế bào. Các enzyme cắt hạn chế là một loại endonuclease của mỗi dòng vi khuẩn (xem chương 9), không cắt DNA của chúng vì đã được methyl hóa ở những điểm cần thiết mà chỉ cắt DNA ngoại lai (của bacteriophage) do chúng không được methyl hóa ở những vị trí nhất định.

Các cơ chế chủ yếu biến đổi của DNA của nó một cách đặc hiệu và cắt hạn chế (restriction) các phân tử DNA không được đánh dấu. Hệ thống R-M ngăn chặn DNA ngoại lai không có hoạt động chức năng nhưng có thể thực hiện tái tổ hợp.

Hệ thống R-M có hai tính chất căn bản:

- Hoạt tính restriction (cắt hạn chế) đặc hiệu chống sự xâm nhập của DNA ngoại lai.

- Hoạt tính bảo vệ DNA của bản thân nó.

Hệ thống R-M lần đầu tiên được phát hiện khi nghiên cứu hiện tượng nhiễm bacteriophage vào *E. coli*. Sự biến đổi thường là methyl hóa nhóm 6-NH_2 của adenine ở những trình tự nucleotide đặc hiệu trong hệ thống kiểu II, C5 của cytosine được methyl hóa. Cắt hạn chế được thực hiện do các hệ thống endonuclease. Chúng nhận biết các điểm đặc hiệu ít nhất ở một sợi DNA không bị biến đổi. Endonuclease cắt đứt sợi ở nhiều điểm nhận biết và sau đó các đoạn ngắn được cắt bằng nucleotide khác.

Phần lớn các DNA bị biến đổi có các đặc tính di truyền không ổn định. Sự methyl hóa thường mất qua tái bản trong tế bào chủ.

Các nghiên cứu cho thấy ở chủng *E. coli* B có 3 gen liên kết chặt với nhau mã hóa cho 3 sản phẩm khuếch tán: các polypeptide phục vụ cho sự cắt hạn chế, cho sự biến đổi và tạo sự đặc hiệu của điểm nhận biết.

DNA sợi kép nhạy cảm hơn với sự biến đổi và cắt hạn chế. Các bacteriophage sợi đơn như M13 và ϕ X 174 ít bị tác động hơn. Sự biến đổi của một sợi DNA đủ để miễn nhiễm DNA lạ. Do vậy, DNA tái bản bán bảo tồn được bảo vệ bởi methyl hóa sợi khuôn.

1. Các methylase của R-M

Methylase là enzyme có hoạt tính xúc tác gắn một nhóm methyl vào một phân tử, ví dụ: DNA methyltransferase có chức năng methyl hóa các gốc C trong DNA.

Enzyme methylase tạo ra một cặp adenine methyl hóa (hoặc cytosine) có vị trí đối xứng chéo nhau. Các methylase có liên quan chặt chẽ với các kiểu cắt hạn chế.

- **Kiểu I.** Các RE có thể tiến hành cả hai sự biến đổi và cắt hạn chế. Một RE kiểu I bao gồm ba tiểu đơn vị khác nhau (R, M và S) chịu trách nhiệm tương ứng cho việc cắt hạn chế, biến đổi và nhận biết trình tự. Ví dụ: *Eco*B và *Eco*K (từ *E. coli* chủng B và K), các trình tự nhận biết của hai enzyme này là TGA$(N)_8$TGCT đối với *Eco*B và AAC$(N)_6$GTGC đối với *Eco*K (trong đó N là một trong bốn loại nucleotide). Nếu vị trí nhận biết không được methyl hóa trong một sợi, thì enzyme RE sẽ methyl hóa sợi đó, nhưng nếu cả hai sợi đều không được methyl hóa thì enzyme RE sẽ cắt DNA (cần có ATP). Phản ứng cắt xuất hiện ở một vị trí không đặc hiệu >1000 bp tính từ trình tự nhận biết; rõ ràng là enzyme duy trì liên kết ở vị trí nhận biết của nó, và DNA được di chuyển qua một vị trí thứ hai trên enzyme hướng tới một vị trí cắt hạn chế.

- Kiểu II. Các RE là dạng chung nhất trong vi khuẩn; chúng chỉ thực hiện phản ứng cắt hạn chế-còn sự biến đổi được tiến hành bởi một enzyme riêng biệt. Trình tự nhận biết cho hầu hết các enzyme này là các trình tự palindrome ngắn (ví dụ: *Eco*RI nhận biết GAATTC) và thực hiện phản ứng cắt (không cần ATP) ở trong hoặc gần kề với vị trí nhận biết. Một vài RE kiểu II cắt cả hai sợi ở cùng vị trí để tạo ra đầu bằng , trong khi các RE khác lại tạo ra đầu dính. Các RE kiểu II là công cụ hữu ích trong công nghệ DNA tái tổ hợp.

- Kiểu III. Các RE này tương tự các enzyme kiểu I. Mỗi RE kiểu III có hai tiểu đơn vị; nó nhận biết một trình tự đối xứng ngắn, và tạo ra một đầu so le (cần có ATP) ở vị trí khoảng 24-26 bp so với vị trí nhận biết. Ví dụ enzyme *Eco*PI, được mã hóa bởi prophage của bacteriophage P1.

2. Adenine và cytosine methyltransferase ở E. coli

Sự methyl hóa A và C thực hiện rộng rãi ở tất cả các loại DNA. Ở prokaryote (kể các bacteriophage và plasmid), vai trò thể hiện rõ là phân hủy DNA ngoại lai. Vai trò phụ là phân biệt DNA bố mẹ với sợi con mới được tổng hợp trong phục hồi các chỗ bắt cặp sai. Ở prokaryote, các sinh vật không có hệ thống cắt hạn chế, vai trò chủ yếu của methyl hóa có lẽ ở sự điều hòa biểu hiện của gen.

Bên cạnh các methylase của hệ thống R-M, còn có các methyltransferase khác ở *E. coli*. Các nucleotide được methyl hóa không gắn trực tiếp vào DNA. Các methylase

chuyển nhóm methyl từ S-adenosine methionine sang adenine và cytosine ở những điểm đặc hiệu trên DNA.

3. Sự methyl hóa DNA của eukaryote

Sự methyl DNA của eukaryote có nhiều điểm khác với ở prokaryote:

- Base bị biến đổi chủ yếu ở eukaryote là 5-methylcytosine. Ở động vật có vú, chỉ có 5-methylcytosine là base bị biến đổi, trong đó 90% là trình tự đối xứng CpG.

- Do chưa hiểu biết rõ hệ thống R-M ở eukaryote nên người ta cho rằng có lẽ methyl hóa không góp phần bảo vệ chống sự xâm nhập của DNA ngoại lai.

Sự methyl hóa của DNA động vật có vú góp phần vào sự điều hòa biểu hiện của gen và biệt hóa tế bào. Các nghiên cứu cho thấy có sự tương quan thuận giữa methyl hóa thấp (hyphomethylation) với hoạt tính của gen và ngược lại giữa sự methyl hóa mạnh với bất hoạt của gen. Sự vắng mặt của các nhóm methyl trên DNA của ruồi giấm chứng tỏ rằng sinh vật đã biệt hóa cao có thể hoạt động không cần cơ chế methyl hóa.

Sự methyl hóa ở động vật có vú có thể đóng vai trò:

- Thay đổi trạng thái của DNA và ức chế ái lực của repressor với operator.

- Sự methyl hóa được truyền đạt theo dòng tế bào soma.

- Methyl hóa có thể tạo sự khác nhau giữa các mô.

- 5-azacytidine, chất đồng đẳng với cytidine, tác động làm tế bào nguyên bào sợi (fibroblast) biệt hóa thành tế bào cơ, có lẽ do cơ chế ức chế methyl hóa DNA và như vậy hoạt hóa gen tương ứng.

Tài liệu tham khảo/đọc thêm

1. Phạm Thành Hổ. 2003. Di truyền học. *NXB Giáo dục,* Hà Nội.

2. Võ Thị Thương Lan. 2002. Sinh học phân tử. *NXB Đại học Quốc gia Hà Nội*, Hà Nội.

3. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc*. New York, USA.

4. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

5. Watson JD, Hopkins NH, Roberts JW and Weiner AM. 2004. Molecular Biology of the Gene. *The Benjamin/Cummings Publishing Company, Inc*. California, USA.

6. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company*, New York, USA.

Chương 8

Điều hòa biểu hiện gen

Như chúng ta đã biết ba quá trình thiết yếu cho sự tồn tại của tế bào, đó là: tái bản, phiên mã và dịch mã. Tuy nhiên, tế bào không thể tồn tại độc lập với môi trường chung quanh. Như vậy, sẽ nảy sinh một vấn đề quan trọng: tế bào sẽ điều chỉnh hoạt động của mình như thế nào cho phù hợp với các biến đổi của môi trường bên ngoài để có thể tồn tại thích ứng? Chương này sẽ đề cập đến các phương thức điều chỉnh đó, tức là các cơ chế điều hòa sự biểu hiện của gen ở các sinh vật prokaryote và eukaryote.

Sự biểu hiện của các gen chịu sự kiểm soát của các cơ chế điều hòa. Các cơ chế này giữ vai trò rất quan trọng cho các hoạt động sống, đáp lại những biến đổi của môi trường bên trong và bên ngoài cơ thể. Biểu hiện gen của các tế bào prokaryote và eukaryote cũng có sự khác nhau đáng kể. Việc điều hòa được thực hiện ở nhiều mức độ khác nhau và liên quan đến từng giai đoạn phát triển. Theo quan niệm về operon, các gen điều hòa (regulatory gene) giữ vai trò quan trọng trong việc đóng và mở các gen cấu trúc (structural gene) để có thể biểu hiện tổng hợp protein đúng lúc, đúng nơi theo nhu cầu cụ thể của tế bào.

Trong mọi tế bào, tất cả các gen đều không hoạt động đồng thời. Ví dụ: tế bào *E. coli* có khoảng 10^7 phân tử protein gồm 3.000 loại khác nhau. Nhiều loại protein có đến 500.000 phân tử, tuy nhiên một số loại khác chỉ khoảng 10 phân tử. Như vậy, không phải loại protein nào cũng được tổng hợp với số lượng lớn như nhau và tế bào phải có những cơ chất để tổng hợp protein một cách tiết kiệm và hợp lý nhất.

Một số gen hoạt động thường xuyên cung cấp sản phẩm liên tục, một số khác chỉ biểu hiện ở những giai đoạn nhất định trong chu trình sống và có thể chỉ hoạt động trong điều kiện môi trường không bình thường. Một số protein cần được tổng hợp với số lượng lớn, một số khác chỉ cần có một phân tử. Do vậy, hoạt tính của gen được điều hòa bởi nhiều cơ chế khác nhau để có hiệu quả tốt nhất trong việc sử dụng nguồn năng lượng của tế bào.

I. Các hiện tượng điều hòa

Để duy trì nội cân bằng (homeostasis) và sự phát triển của cơ thể, các sinh vật đã có các cơ chế điều hòa khác nhau. Các kiểu điều hòa đều bắt nguồn từ sự biểu hiện của các gen.

1. Điều hòa thích nghi

Một số amip (ameba) biểu hiện sự thay đổi hình thái và sinh lý đặc biệt để đáp lại các điều kiện môi trường khác nhau. Khi các amip được cho vào nước, chúng chuyển từ

dạng amip sang dạng có lông để bơi. Khi môi trường thiếu dinh dưỡng chúng có thể chuyển thành các dạng tương tự như biểu bì.

Vi khuẩn trong môi trường dinh dưỡng tối thiểu có khả năng tổng hợp amino acid. Nhưng khi bổ sung amino acid vào môi trường nuôi, vi khuẩn sẽ ngừng tổng hợp amino acid. Lúc nguồn amino acid từ ngoài bổ sung vào đã hết, tế bào vi khuẩn lại tự tổng hợp lại amino acid cho bản thân.

Các biến đổi nêu trên là thuận nghịch, chứng tỏ sự thay đổi chức năng ở đây không phải do biến dị di truyền. Các hiện tượng trên còn cho thấy việc xuất hiện hay biến mất các cấu trúc mới không làm ảnh hưởng đến tiềm năng di truyền sẵn có. Có thể cho rằng, có trường hợp một số gen hoạt động, nhưng cũng có trường hợp một số gen ngừng biểu hiện. Các hiện tượng được đề cập trên đều do cơ chế điều hòa thích nghi (adaptive regulation) chi phối.

2. Hoạt động nối tiếp của các gen

Khi bacteriophage xâm nhiễm vi khuẩn, DNA của nó lúc đầu sẽ tái bản, sau đó các protein khác nhau mới được tổng hợp nên để tạo thành vỏ. Như vậy, có các gen "sớm" tạo ra enzyme tái bản DNA và các gen "muộn" xác định các thành phần vỏ protein. Điều đó chứng tỏ có cơ chế điều hòa chức năng của gen diễn ra theo một trình tự nghiêm ngặt. Đây là kiểu điều hòa nối tiếp (sequential regulation). Hoạt động nối tiếp của các gen còn thể hiện rõ trong quá trình phát triển cá thể của các sinh vật eukaryote đa bào.

3. Biệt hóa tế bào

Nhiều sinh vật bậc cao như con người chứa nhiều tỷ tế bào bắt nguồn từ một hợp tử do phân chia nguyên nhiễm. Từ một hợp tử ban đầu đến khi trưởng thành, cơ thể người có khoảng 200 loại tế bào khác nhau. Mỗi loại tế bào chỉ biểu hiện một phần thông tin của mình. Quá trình chuyên môn hóa chức năng của tế bào được gọi là sự biệt hóa hay phân hóa (differentiation).

Tuy có sự biệt hóa, nhưng tế bào vẫn giữ nguyên vẹn khả năng di truyền của mình. Một ví dụ rất rõ là nuôi cấy mô tế bào thực vật (plant tisue and cell culture): người ta có thể nuôi cấy một phần mô phân sinh trong môi trường dinh dưỡng tổng hợp cho đến khi chúng phát triển thành cây *in vitro* hoàn chỉnh (plantlet), các cây này sau đó được đưa ra trồng trong điều kiện tự nhiên và đã ra hoa kết quả.

4. Khái quát về điều hòa ở prokaryote và eukaryote

Có sự khác nhau đáng kể giữa prokaryote và eukaryote trong điều hòa biểu hiện của gen. Các tế bào eukaryote có cấu tạo phức tạp hơn nhiều nên cơ chế điều hòa cũng phức tạp hơn prokaryote.

Ở prokaryote, mục đích của sự điều hòa biểu hiện gen là nhằm điều chỉnh hệ enzyme cho phù hợp với các tác nhân dinh dưỡng và lý hóa của môi trường, đảm bảo

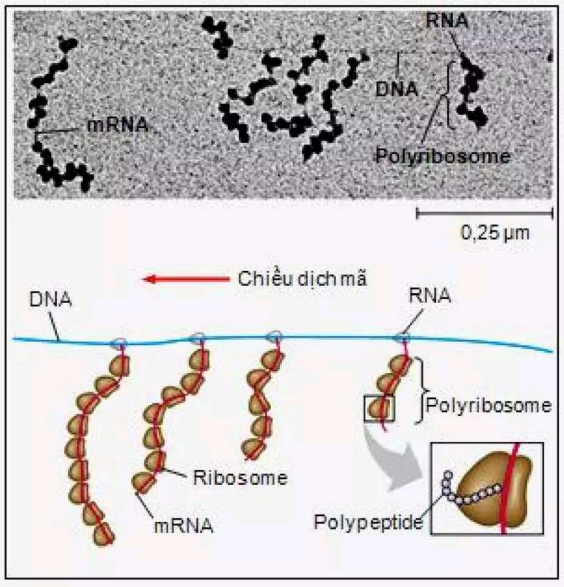
được hai yêu cầu chính của tế bào là sinh trưởng và sinh sản. Sự điều hòa ở đây rất linh động và có tính thuận nghịch. Ở eukaryote, do tế bào không tiếp xúc trực tiếp với môi trường, nên sự điều hòa ở đây không còn nhằm mục đích đối phó với các biến động ở ngoại bào. Sự điều hòa ở eukaryote hướng đến việc chuyên biệt hóa từng loại tế bào vào từng cấu trúc và chức năng riêng và vì thế không mang tính thuận nghịch.

Ba thành phần chính của sự điều hòa biểu hiện gen là: 1) Tín hiệu gây ra đáp ứng làm thay đổi biểu hiện gen; 2) Giai đoạn được thực hiện sự điều hòa trong quá trình từ tái bản đến dịch mã; và 3) Cơ chế phân tử của sự điều hòa biểu hiện gen.

4.1. Sự biểu hiện của gen ở prokaryote

Bộ máy di truyền của sinh vật prokaryote là một DNA mạch vòng chứa một số lượng gen giới hạn được phiên mã ở trạng thái tiếp xúc trực tiếp với tế bào chất (Hình 8.1).

Chu trình tế bào ngắn và không có sự biệt hóa tế bào. Vì thế, hoạt động của các gen được điều hòa do các nhu cầu của tế bào khi cần thiết. Tác động của các nhân tố môi trường làm những gen tương ứng được mở để phiên mã, dịch mã tổng hợp protein hay có hiệu quả ngược làm dừng lại.



Hình 8.1. Sự biểu hiện gen ở prokaryote

4.2. Sự biểu hiện của gen ở eukaryote

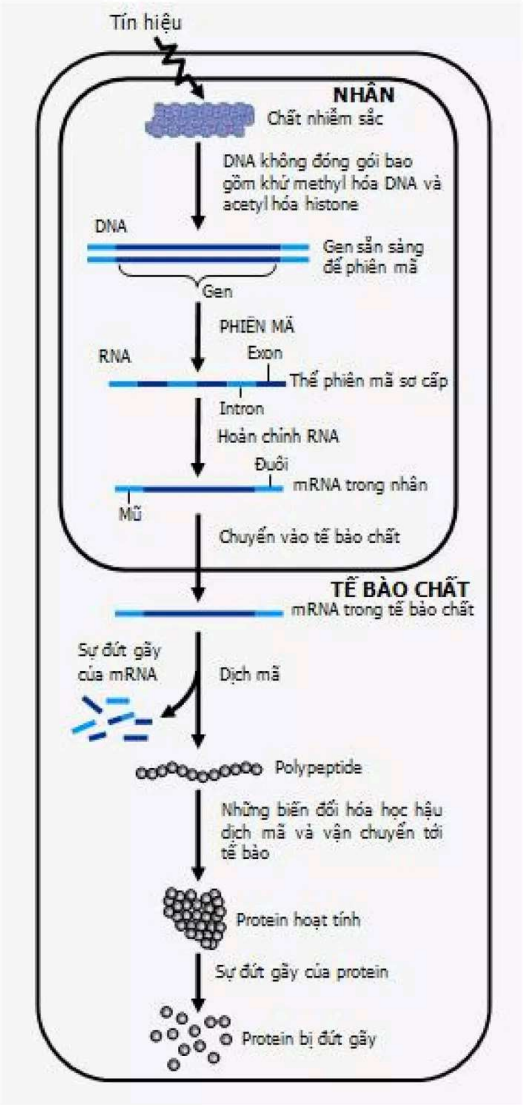
Khác với prokaryote, nhiễm sắc thể của eukaryote có cấu trúc phức tạp. Ngay trên cấu trúc nhiễm sắc thể có sự tham gia của các protein histone có vai trò điều hòa biểu

hiện của gen. Sự điều hòa biểu hiện gen ở eukaryote phải qua nhiều mức điều hòa phức tạp hơn so với prokaryote và qua nhiều giai đoạn như: nhiễm sắc thể tháo xoắn, phiên mã, biến đổi hậu phiên mã, mRNA rời nhân ra tế bào chất, dịch mã và biến đổi hậu dịch mã (Hình 8.2).

Ngoài ra, đa số eukaryote có cơ thể đa bào và mỗi tế bào có biểu hiện sống không phải tự do, mà chịu sự biệt hóa theo các chức năng chuyên biệt trong mối quan hệ hài hòa với cơ thể.

Các vi khuẩn thường phản ứng trực tiếp với môi trường và biểu hiện gen thuận nghịch, như có đường lactose thì mở operon để phân hủy, khi hết đường thì operon đóng lại. Trong khi đó, các tế bào eukaryote có những con đường biệt hóa khác nhau và sự chuyên hóa là ổn định thường xuyên trong đời sống cá thể. Ngoài sự biệt hóa tế bào, các cơ thể eukaryote đa bào còn trải qua quá trình phát triển cá thể với nhiều giai đoạn phức tạp nối tiếp nhau, trong đó có những gen chỉ biểu hiện ở phôi và sau đó thì dừng hẳn.

Tất cả những điểm nêu trên cho thấy sự điều hòa biểu hiện của gen eukaryote phức tạp hơn nhiều, mà hiện nay lại được biết ít hơn prokaryote.



Hình 8.2. Sự biểu hiện gen ở eukaryote

II. Các mức độ điều hòa

Các cơ chế điều hòa sự biểu hiện của gen có thể tác động ở một hay nhiều mức độ khác nhau. Sự điều hòa có thể xảy ra ở mức độ gen bằng sự kiểm soát thời gian và tốc độ phiên mã. Các cơ chế khác có thể hoạt động lúc dịch mã hoặc sau dịch mã.

1. Mức độ chất nhiễm sắc

Ngay trên chất nhiễm sắc có thể thực hiện các kiểu sau:

- DNase cắt một số vùng trên genome làm tháo xoắn để các gen biểu hiện. Hai vùng được lưu ý đó là các vùng nhạy cảm (sensible) và siêu nhạy cảm (hypersensible).

- Các vùng nhạy cảm có liên quan đến các gen có hoạt tính cao và những gen đã qua biểu hiện rồi (như các gen hoạt động ở phôi). Các vùng siêu nhạy cảm liên quan đến các gen có hoạt tính rất cao (như các gen histone).

- DNA Z (DNA trái) là dạng cấu trúc siêu xoắn có thể liên quan đến đóng mở gen.

- Methyl hóa các base. Ở các prokaryote sự methyl hóa có thể thực hiện đối với A và C, còn ở eukaryote sự methyl hóa chỉ thực hiện với C vị trí thứ 5. Methyl hóa làm gen ngừng hoạt động. Ví dụ: nhiễm sắc thể X bất hoạt ở người thuộc loại siêu methyl hóa. Nói chung, sự thay đổi cấu hình (reconfiguration) có thể ảnh hưởng đến sự biểu hiện của gen.

2. Mức độ phiên mã

Đây là sự điều hòa ảnh hưởng trực tiếp đến việc mở hoặc đóng của gen. Kiểu điều hòa này thường gặp trong điều hòa trao đổi chất, cũng như các quá trình biệt hóa tế bào.

- Sự tác động của các trình tự *cis* (gần kề, liền kề) nằm trên cùng mạch DNA như enhancer (vùng tăng cường) làm tăng sự phiên mã.

- Điều hòa bởi các nhân tố *trans* (cách quãng, từ xa) do các nhân tố không nằm cùng trên một mạch DNA.

- Chọn lựa promoter thích hợp.

- Sự suy yếu/suy thoái.

3. Mức độ hậu phiên mã

Sự điều hòa có thể biểu hiện ở mức tác động lên mRNA, chúng ta đã gặp trường hợp trên khi mRNA bị cắt bỏ các intron và gắn các exon lại với nhau để tạo thành mRNA hoàn chỉnh (RNA processing). Như vậy, các hệ thống ảnh hưởng đến sự hoàn chỉnh của mRNA có thể kiểm tra gián tiếp biểu hiện của gen tương ứng. Các mRNA của eukaryote còn có những đoạn không mã hóa liên quan tới thời gian tồn tại và ra khỏi nhân vào tế bào chất.

- Splicing khác nhau.

- Điểm polyadenine hóa khác nhau (polyadenylation).

- Đột biến trên phân tử mRNA.

- Bán chu kỳ phân hủy của mRNA.

- Sự bảo tồn các RNA trong tế bào.

4. Mức độ dịch mã

Sự biến đổi của các nhân tố khởi đầu IF (inititation factor). Là các protein kết hợp với tiểu đơn vị của ribosome vào giai đoạn khởi động của quá trình dịch mã.

5. Mức độ hậu dịch mã

Ở đây có sự điều hòa hoạt tính của protein. Sau khi mạch polypeptide được tổng hợp, các protein nhiều khi phải trải qua các biến đổi thứ cấp trước khi biểu hiện hoạt tính (chức năng). Ví dụ: trypsin là enzyme phân giải protein trong dạ dày chỉ có được hoạt tính sau khi chất tiền thân của nó (pro-enzyme không có hoạt tính) bị cắt mất một đoạn polypeptide.

Các protein có thể chịu những biến đổi lập thể như sự kết hợp các enzyme với một số sản phẩm đặc biệt có thể làm thay đổi cấu trúc không gian của chúng dẫn đến mất hoạt tính.

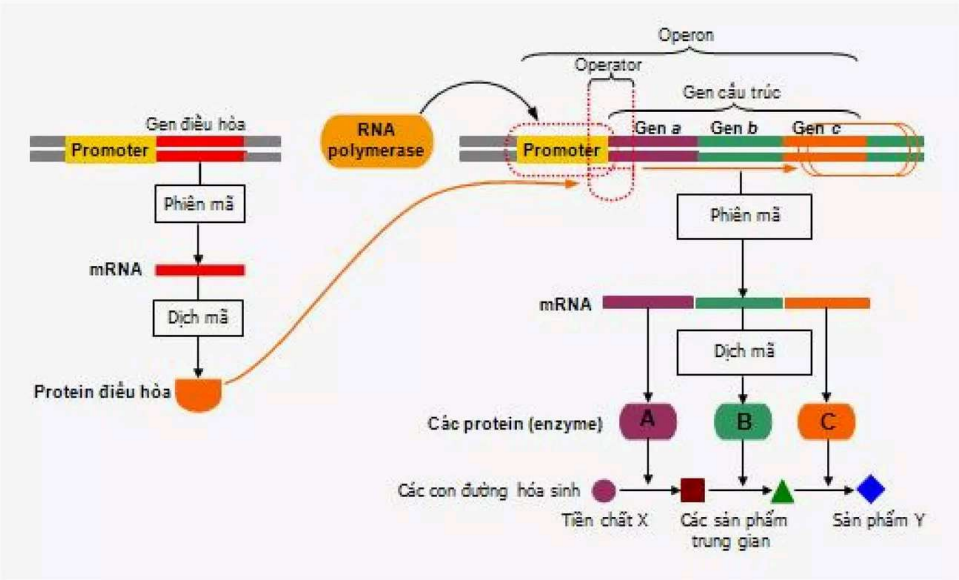
- Các quá trình glycosylation, phosphorylation… tức là gắn thêm các nhóm chất như đường, phosphor… để protein có hoạt tính/chức năng sinh học.

- Peptide tín hiệu là đoạn gồm khoảng 20 amino acid nằm gần phía đầu N của polypeptide, có vai trò gắn polypeptide và ribosome đang tổng hợp mạch này với mạng lưới nội sinh chất. Trong bộ máy Golgi, polypeptide được phóng thích ra ngoài.

- Sự phóng thích ra protein có chức năng sinh học từ một phức hợp, như từ pro-insulin thành insulin.

III. Điều hòa biểu hiện gen ở prokaryote

Các gen được phiên mã tạo RNA, được gọi là các gen cấu trúc. Các protein được dịch mã từ mRNA có thể là enzyme hoặc không phải enzyme. Trong số các protein không phải enzyme có các protein điều hòa (regulatory protein), chúng tương tác với các trình tự DNA đặc hiệu để kiểm soát hoạt tính phiên mã của các gen cấu trúc. Các gen tổng hợp các protein điều hòa được gọi là các gen điều hòa (regulatory gen). Phía trước mỗi gen cấu trúc (hoặc một nhóm gen) có một trình tự promoter, nơi RNA polymerase nhận biết (Hình 8.3). Cơ chế điều hòa ở prokaryote chủ yếu được thực hiện thông qua operon. Đây là khái niệm chỉ tồn tại ở prokaryote.



Hình 8.3. Phương thức chung điều hòa biểu hiện gen ở prokaryote

1. Cấu trúc của promoter

Thực chất của khởi sự phiên mã là quan hệ trực tiếp giữa RNA polymerase và promoter. Khi RNA polymerase gắn vào promoter, nó sẽ phiên mã tạo phân tử RNA.

Phần lớn promoter ở *E. coli* về căn bản có cùng cấu trúc:

Nếu base đầu tiên được phiên mã thành mRNA (luôn là purine, thường là adenine) được đánh số +1, thì tất cả các base phía 5' hay "phía trước" so với nó không được phiên mã là số trừ (–). Ngay phía trước +1 có 6 base thường với trình tự TATAAT ở xung quanh –10, và trình tự TTGACA (trình tự liên ứng-consensus sequence) ở xung quanh –35. Cả hai trình tự phối hợp nhau cho phép RNA polymerase gắn vào và khởi sự dịch mã, trình tự –35 tạo điều kiện đầu tiên cho việc gắn vào.

2. Cấu trúc của operon

Operon là đơn vị phiên mã gồm ít nhất một promoter và mRNA ở bước tiếp theo để mã hóa cho các trình tự của một hay nhiều chuỗi polypeptide. Tuy nhiên, operon có thể có một hay nhiều điểm điều hòa khác với promoter. Các gen không chịu sự điều hòa do tác động môi trường, tạo sản phẩm thường xuyên, được gọi là các gen cấu trúc. Số lượng sản phẩm của các gen này có thể dao động phụ thuộc vào ái lực tương đối của các

promoter của chúng đối với RNA polymerase. Các promoter có ái lực mạnh (strong promoter) tạo ra nhiều sản phẩm của gen hơn các promoter có ái lực yếu. Các gen mà sản phẩm protein của chúng được tổng hợp đáp lại với các nhân tố môi trường, thường được điều khiển bởi một hay nhiều protein điều hòa. Trình tự DNA bên trong operon, nơi mà protein ức chế gắn vào, được gọi là operator (điểm điều hành). Việc gắn protein ức chế lên operator ngăn cản sự phiên mã của tất cả các gen cấu trúc trên cùng một operon. Sự kiểm soát như vậy đối với với gen gọi là kiểm soát âm. Các operon của vi khuẩn thường tạo ra các mRNA đa gen, nhưng mRNA của eukaryote chỉ một gen.

Các protein cần thiết cho biểu hiện gen được gọi là chất hoạt hóa. Chúng có thể gắn với các điểm khởi sự nằm bên trong của promoter của operon hay điểm tăng cường hoặc có thể gắn ở những trình tự xa operon. Việc gắn của protein điều hòa vào điểm khởi đầu (initiator) hay enhancer, kích thích sự phiên mã của các gen cấu trúc, được gọi là cơ chế kiểm soát dương. Sự kích thích để các gen điều hòa phản ứng có thể là từ các phân tử tương đối nhỏ như đường, amino acid đến các phân tử lớn hơn như các phức hợp hormone steroid và các protein thụ thể (receptor). Chất làm cho gen phiên mã được gọi là chất cảm ứng, có tác động ngược với chất kìm hãm. Các gen cảm ứng thường tham gia vào các phản ứng thoái dưỡng (catabolic reaction), như phân hủy các polysaccharide thành đường đơn. Các gen ức chế thường tham gia vào các phản ứng biến dưỡng thực hiện việc tổng hợp các chất như amino acid từ các tiền chất đơn giản hơn.

3. Điều hòa thoái dưỡng: Kiểm soát âm-cảm ứng

Trong thoái dưỡng, các chất thức ăn được phân hủy dễ dàng tạo năng lượng hoặc các chất cần thiết cho quá trình tổng hợp. Cơ chế điều hòa ở đây là sự có mặt của cơ chất (ví dụ lactose) dẫn tới tổng hợp các enzyme phân hủy.

Ví dụ điển hình cho trường hợp này là operon lactose của *E. coli*. β-galactosidase là enzyme có chức năng đôi. Chức năng đầu tiên của nó là thoái dưỡng lactose thành glucose và galactose. Chức năng thứ hai của nó là chuyển liên kết 1-4 của glucose và galactose thành liên kết 1-5 của allolactose. Bình thường enzyme này không hiện diện ở nồng độ cao trong tế bào, khi vắng mặt lactose trong môi trường. Ngay sau khi cho lactose vào môi trường nuôi khi không có glucose, enzyme này bắt đầu được tạo ra. Sự vận chuyển lactose xuyên qua màng tế bào có hiệu quả nhờ protein vận chuyển galactoside permease. Protein cũng xuất hiện với nồng độ cao khi có lactose trong môi trường.

Sự điều hòa của operon lactose còn phụ thuộc vào nồng độ glucose trong môi trường. Nồng độ glucose này lại kiểm soát nồng độ bên trong tế bào của phân tử nhỏ cAMP (cyclic adenosine monophosphate), là chất bắt nguồn từ ATP và làm tín hiệu báo động cho tế bào. Tế bào có xu hướng sử dụng glucose hơn là lactose để làm nguồn carbon vì glucose được biến dưỡng trực tiếp cung cấp carbon và tạo năng lượng. Các enzyme biến dưỡng glucose thuộc loại cấu trúc và tế bào tăng trưởng tối đa với nguồn

glucose. Khi nguồn glucose cạn, tế bào phản ứng lại bằng cách tạo ra c-AMP. Việc tăng nồng độ c-AMP trong tế bào gây nên hàng loạt sự kiện, trong sự hiện diện của lactose, dẫn đến sự phiên mã các gen cấu trúc của operon lactose.

3.1. Cấu trúc của operon lactose

Hệ thống lactose (lactose system) bình thường (Hình 8.4) gồm có gen điều hòa (i hoặc R) và operon mang trình tự promoter (P) locus operator (O) và ba gen cấu trúc cho β-galactosidase (Z), permease (Y) và transacetylase (A). Nhiều đột biến ở các locus này đã được phát hiện.

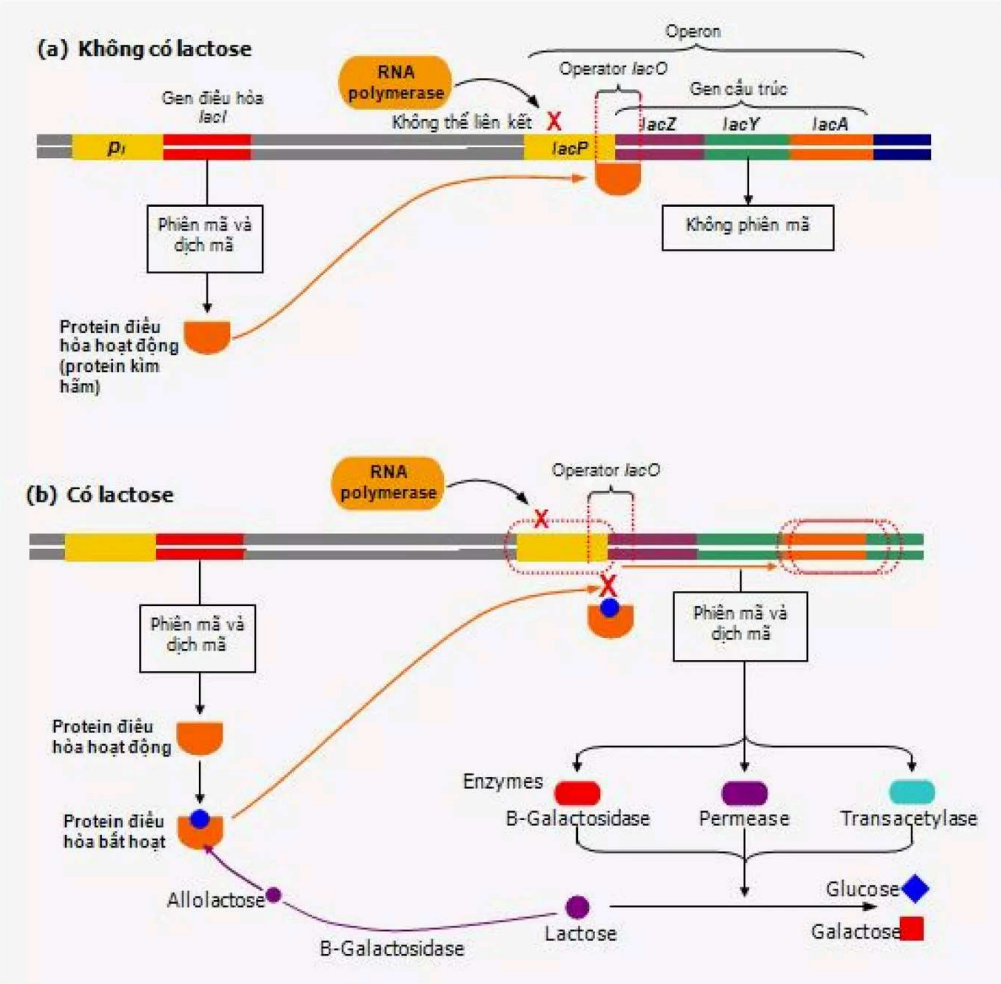
3.2. Hoạt động của hệ thống

- Điều kiện cảm ứng (có lactose). Lactose được chuyển vào tế bào rất yếu vì chỉ có vài phân tử permease làm việc. Khi vào trong tế bào, một số lactose (liên kết β-1,4) được chuyển thành allolactose (liên kết β-1,6) nhờ β-galactosidase. Allolactose là chất cảm ứng, nó gắn vào protein kìm hãm và gây biến đổi cấu hình tạo phức hợp allolactose-repressor. Phức hợp này mất khả năng gắn operator. Lúc này operon được mở, RNA polymerase bắt đầu phiên mã các gen cấu trúc. Toàn bộ sự kiện diễn ra như trên hình 8.4b.

- Điều kiện không cảm ứng (không có lactose). Gen điều hòa của operon thường xuyên tổng hợp protein kìm hãm (repressor protein) ở mức độ thấp, vì nó có promoter ít hiệu quả. Sự tổng hợp các protein này bị tác động do nồng độ lactose trong tế bào. Ngược lại, promoter bình thường của operon lac gắn với RNA polymerase rất có hiệu quả. Khi không có đường lactose, protein điều hòa hoạt động (active regulator protein) còn gọi là protein kìm hãm gắn vào promoter hay "đọc" trình tự operator vì protein kìm hãm chiếm đoạn này. Như vậy, sự phiên mã của tất cả các gen cấu trúc của operon lac bị dừng (Hình 8.4a).

Do số lượng permease tăng, nên lactose vào tế bào với số lượng lớn và được phân hủy bởi β-galactosidase. Khi lactose được sử dụng hết, các protein kìm hãm gắn trở lại vào operator làm operon bị đóng; sự phiên mã các gen cấu trúc bị dừng.

Bản thân gen điều hòa *lacI* chỉ có một promoter (P_i) và gen cấu trúc của protein kìm hãm. Promoter này yếu, khi các protein kìm hãm có số lượng lớn, nó bị các protein này gắn vào làm dừng phiên mã.



Hình 8.4. Operon lactose và hoạt động của nó

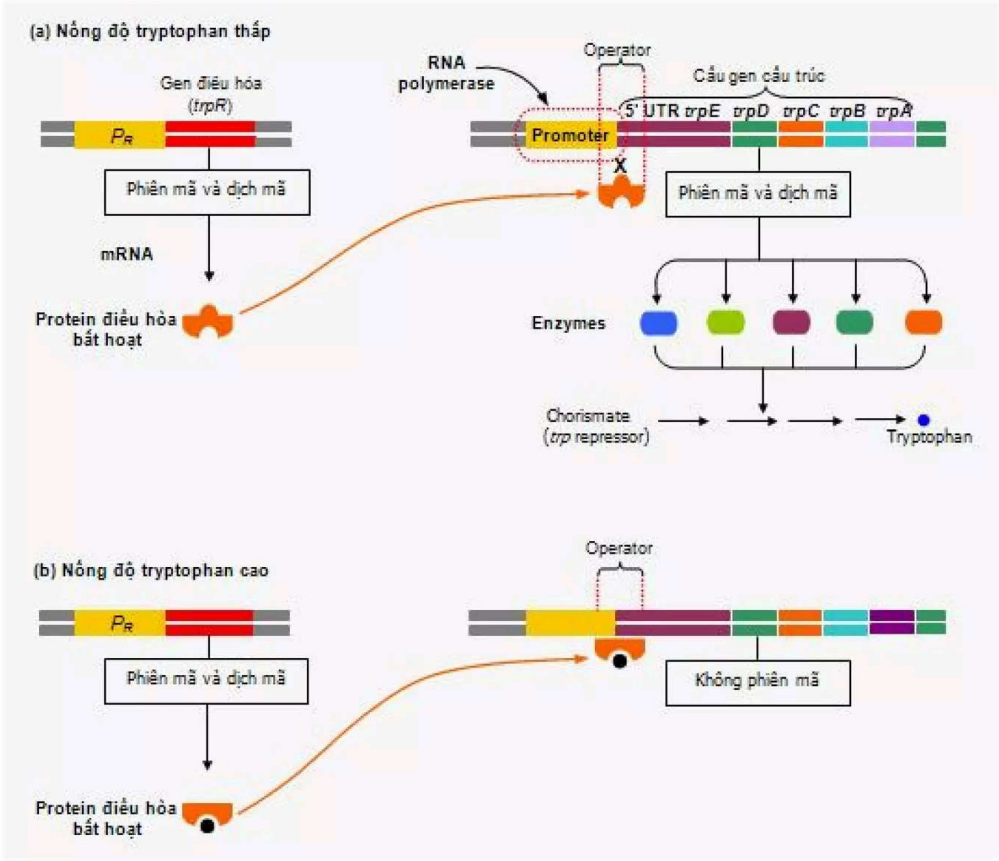
4. Điều hòa biến dưỡng: Kiểm soát âm-ức chế

Biến dưỡng (anabolism) là quá trình tổng hợp nên các chất cần thiết cho tế bào. Ví dụ tổng hợp các amino acid.

Quá trình tổng hợp tryptophan bắt đầu từ tiền chất tryptophan là chorismic acid, trải qua 5 giai đoạn kế tiếp do enzyme xúc tác. Hệ thống tổng hợp amino acid tryptophan ở *E. coli* là ví dụ điển hình về operon bị kìm hãm do sự kiểm soát âm.

4.1. Cấu trúc và hoạt động

Hệ thống tryptophan cũng có cấu trúc tương tự hệ thống lactose gồm gen điều hòa *trpR* và operon tryptophan (promoter, operator và 5 gen cấu trúc). Các gen cấu trúc xác định 5 enzyme được xếp theo thứ tự tương ứng với chức năng xúc tác theo trình tự các phản ứng của chuỗi biến dưỡng tryptophan (Hình 8.5).



Hình 8.5. Operon tryptophan

Sự khác nhau căn bản với hệ thống lactose là ở gen điều hòa. Gen điều hòa của hệ thống tryptophan tổng hợp thường xuyên aporepressor protein, là chất kìm hãm mà riêng nó không có hoạt tính. Khi tryptophan dư thừa nó trở thành chất corepressor (đồng kìm hãm) và kết hợp với aporepressor thành phức hợp kìm hãm (holorepressor) có hoạt tính. Phức hợp này gắn vào operator của operon tryptophan (*trp*) làm dừng phiên mã các gen cấu trúc. Khi nồng độ tryptophan thấp, nó tách khỏi phức hợp kìm hãm và aporepressor mất hoạt tính. Lúc này các operator lại được mở và RNA polymerase dịch mã 5 gen cấu trúc để tổng hợp 5 enzyme tạo tryptophan (Hình 8.5). Sự điều hòa kiểu này còn gọi là điều hòa ức chế ngược (retro-inhibition) do sản phẩm cuối cùng có mối liên hệ ngược (feed-back).

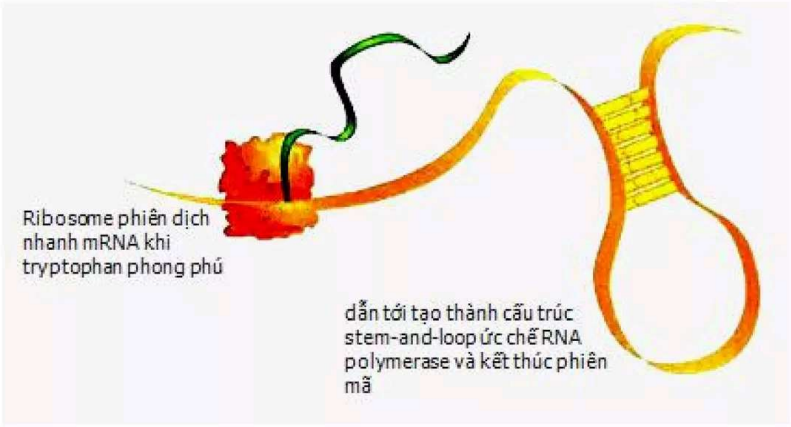
Như vậy, hoạt động của hệ thống này ngược lại với hệ thống lactose: khi có tryptophan thì operon bị đóng, thiếu tryptophan thì gen được mở.

4.2. Sự suy yếu (attenuation)

Kiểu điều hòa thứ hai được phát hiện ở operon tryptophan được gọi là sự suy yếu. Ở đầu 5' của mRNA đa gen (polycistronic) của operon này có 5 enzyme. Đoạn này được gọi là trình tự leader (trình tự chỉ huy). Một phần của trình tự này được phiên mã tạo leader peptide gồm 14 amino acid, mà chức năng đến nay chưa rõ.

Cơ chế kiểm tra mà ở đó sự tổng hợp mRNA được bắt đầu nhưng kết thúc sớm với chiều dài mRNA ngắn hơn được gọi là sự suy yếu. Ví dụ: các tế bào *E. coli* đang tăng trưởng trong môi trường thiếu một amino acid nào đó, nhưng chưa đủ các enzyme cần có cho sự tổng hợp của tất cả 20 amino acid cần thiết để tạo ra protein. Khi thêm một amino acid (tryptophan) vào môi trường sẽ làm giảm đáng kể sự tổng hợp của các enzyme cần cho sự tạo amino acid đó. Phản ứng này có tính thích ứng và duy trì các nguồn enzyme không còn cần nữa khi sản phẩm cuối cùng của chuỗi sinh tổng hợp (tryptophan) đang có trong tế bào (ức chế ngược).

Operon tryptophan có phương thức điều hòa hoạt động gen thứ hai, hoàn toàn độc lập với hệ thống repressor-operator (chất kìm hãm đoạn điều hành). Ở operon tryptophan, sự tổng hợp mRNA bắt đầu từ 161 base trước codon khởi sự của *trpE*, enzyme cấu trúc đầu tiên được phiên mã. Đột biến do mất đoạn DNA ngay phía trước *trpE*, giải phóng *trpE* khỏi sự kìm hãm của phức hợp represser-operator. Các đột biến này làm tăng mức độ biểu hiện của cả operon lên gấp sáu lần. Tinh sạch các DNA được phiên mã *in vitro* cho thấy phần lớn mRNA kết thúc ở ngay base 139 và không bao giờ đạt tới *trpE*. Ở đây, có sự kìm hãm phiên mã ngay trước gen cấu trúc và khi mất nó sự biểu hiện của gen có hiệu quả hơn. Đoạn dài của mRNA nằm trước *trpE* được gọi là leader (*trpL*) và đoạn kìm hãm phiên mã được gọi là attenuator (*trpa*, chứ không phải *trpA*). Cuối cùng, quan sát cho thấy sự suy yếu dao động theo nồng độ của tryptophan thấp, nhiều phân tử mRNA được tạo ra hơn qua attenuator và toàn bộ operon được phiên mã. Sự suy yếu là một phương thức khác để kiểm soát sự biểu hiện của gen.



Hình 8.6. Sự suy yếu xảy ra khi cấu trúc thứ cấp đặc biệt tạo thành trong mRNA. mRNA được trình bày ở đây là từ trình tự leader của operon *trp*. Nếu tryptophan trong tế bào phong phú

và trình tự leader được phiên dịch nhanh, thì mRNA sẽ tạo thành cấu trúc thân và quai (stem-and-loop) hoạt động như là một tín hiệu kết thúc phiên mã.

Cơ chế suy yếu thực hiện sự kết hợp giữa phiên mã và dịch mã. RNA polymerase bắt đầu phiên mã các gen mã hóa cho các enzyme sinh tổng hợp cần thiết để tạo ra tryptophan. Ví dụ: khi hiện diện tryptophan nó nhanh chóng đạt tới điểm (trên RNA mới được tổng hợp) gọi là attenuator và dừng phiên mã tại đây.

Attenuator có chức năng như điểm kết thúc chỉ khi tryptophan hiện diện. Sự phiên mã dừng khi trình tự leader đạt attenuator. Nếu có đủ trytophan trong tế bào, sự dịch mã của các codon được thực hiện, và cả phiên mã và dịch mã được thực hiện hoàn chỉnh cho đến lúc RNA polymerase đạt tới điểm attenuator, nơi kết thúc sự phiên mã. Khi không có trytophan, dịch mã dừng và phần leader của mRNA cuộn lại sao cho attenuator không thực hiện chức năng, trong trường hợp này sự tổng hợp của mRNA toàn vẹn cho operon tryptophan được tạo ra. Với sự hiện diện của các enzyme đó, trytophan được tích lũy với số lượng lớn trong tế bào. Attenuator là dạng tinh tế của sự ức chế, điều hòa hoạt động gen, tạo đặc hiệu cho sự tổng hợp amino acid trong tế bào.

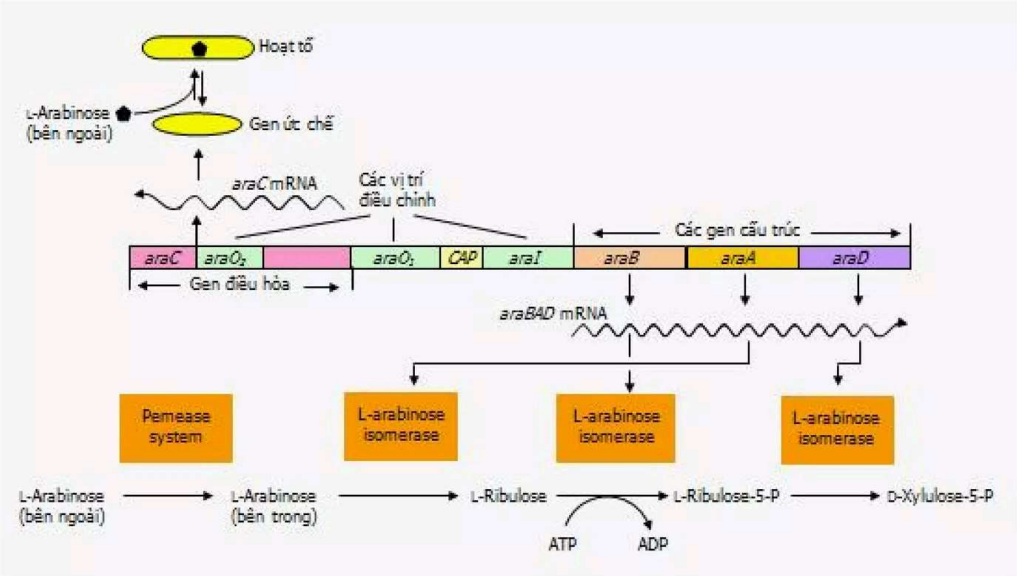
Cơ chế repressor điều hòa thô hệ thống tryptophan, trong khi đó hệ thống cơ chế attenuation kiểm soát nồng độ tryptophan một cách tinh tế. Sự suy yếu của operon *Trp* cũng nhạy cảm với nồng độ của một số amino acid như histidine và leucine.

5. Kiểm soát dương và cảm ứng

Cơ chế điều hòa dương, cảm ứng được tìm thấy ở operon arabinose của *E. coli*. Arabinose là chất đường, cần ba enzyme (được mã hóa bởi các gen *araB, araA* và *araD*) cho sự biến dưỡng. Hai gen khác nằm xa operon góp phần đưa arabinose vào tế bào. Gen điều hòa *araC* nằm gần các gen *B, A* và *D*. Sản phẩm của gen *araC* là protein kìm hãm của operon khi không có đường arabinose.

Tuy nhiên, khi có đường arabinose trong tế bào, nó gắn với repressor (protein araC) hình thành nên phức hợp activator (hoạt tố) tạo thuận tiện cho sự phiên mã của RNA polymerase.

Trên thực tế, các nghiên cứu cho thấy cơ chế điều hòa này rất phức tạp. Ví dụ: cAMP và protein hoạt hóa thoái dưỡng (catabolite gene activator protein) còn được gọi là CRP (cyclic AMP receptor protein-protein thể nhận cAMP) đều tham gia vào sự điều hòa của hệ thống arabinose.



Hình 8.7. Operon arabinose của *E. coli*

IV. Điều hòa hoạt tính của eukaryote

Các cơ chế điều hòa ở eukaryote có thể xảy ra ở 5-6 mức độ khác nhau (xem mục II). Trong phần này chỉ nhấn mạnh thêm một số đặc điểm của điều hòa hoạt động gen ở eukaryote:

- Ở các operon của prokaryote, các gen điều hòa và các promoter thường nằm gần nhau, nhưng ở eukaryote các gen điều hòa ít khi nằm gần các promoter do chúng kiểm soát.

- Các enhancer là những trình tự cùng nằm trên một phân tử với các promoter có thể có hàng trăm cặp base ở phía trước hoặc phía sau promoter mà chúng kích thích.

- Trình tự điều hòa 5' ở phía trước có promoter ở eukaryote thường rất dài, có khi hàng chục kilobase (kb).

- Có nhiều kiểu điều hòa ở dạng các nhân tố có tác động từ xa (*trans*) là các protein.

Sự phiên mã có thể được kích thích bởi các tín hiệu khác nhau. Sự điều hòa hoạt động gen ở prokaryote phần lớn đáp ứng lại các tín hiệu bên trong.

1. Các promoter

Tương tự như ở vi khuẩn, các promoter của eukaryote cùng nằm phía trước điểm xuất phát của mRNA và có những trình tự được bảo tồn trong tiến hóa. Hộp TATA, định hướng cho RNA polymerase bắt đầu phiên mã, nằm khoảng dưới 30 basepair (bp) ở động vật có vú, và 60-120 ở nấm men. Hộp TATA hoạt động có hiệu quả cùng với hai trình tự tương ứng phía trước khoảng 40 bp là CCAAT và 110 bp là trình tự giàu GC.

Sự thay đổi hộp TATA làm giảm tốc độ phiên mã. Hiệu quả của tốc độ phiên mã được đo bằng sự thay đổi của từng base trong promoter, các thay đổi base ngoài hộp TATA và các trình tự phía trước không gây hiệu quả đối với tốc độ phiên mã, trong khi đó các thay đổi ở những yếu tố được nêu trên làm giảm đáng kể tốc độ phiên mã. Khác với promoter của prokaryote, các promoter của eukaryote khó đảm bảo nhận biết các tín hiêu một cách đầy đủ để RNA polymerase khởi sự phiên mã *in vitro*. Hộp TATA và các trình tự phía trước phải được nhận biết bởi các protein điều hòa, và chính các protein này gắn với các điểm nhất định trên chúng và hoạt hóa sự phiên mã.

2. Các enhancer

Enhancer là các trình tự có tác động *cis* (liền kề), chúng tăng tốc độ phiên mã đáng kể từ các promoter nằm ngay trên cùng một phân tử DNA. Tính độc đáo của các enhancer là ở chỗ chúng có khả năng thực hiện một cách hiệu quả khi ở một khoảng cách rất xa, có khi đến vài nghìn cặp base. Ngoài ra, chúng có hoạt động chức năng ở bất kỳ hướng nào, dù ở phía trước hay phía sau promoter.

Xét trên một góc độ nào đó, các enhancer tương tự promoter. Cụ thể, chúng được tổ chức gồm một dãy các trình tự có tác động *cis* để nhận biết các nhân tố tác động từ xa (*trans*) (Hình 8.8).

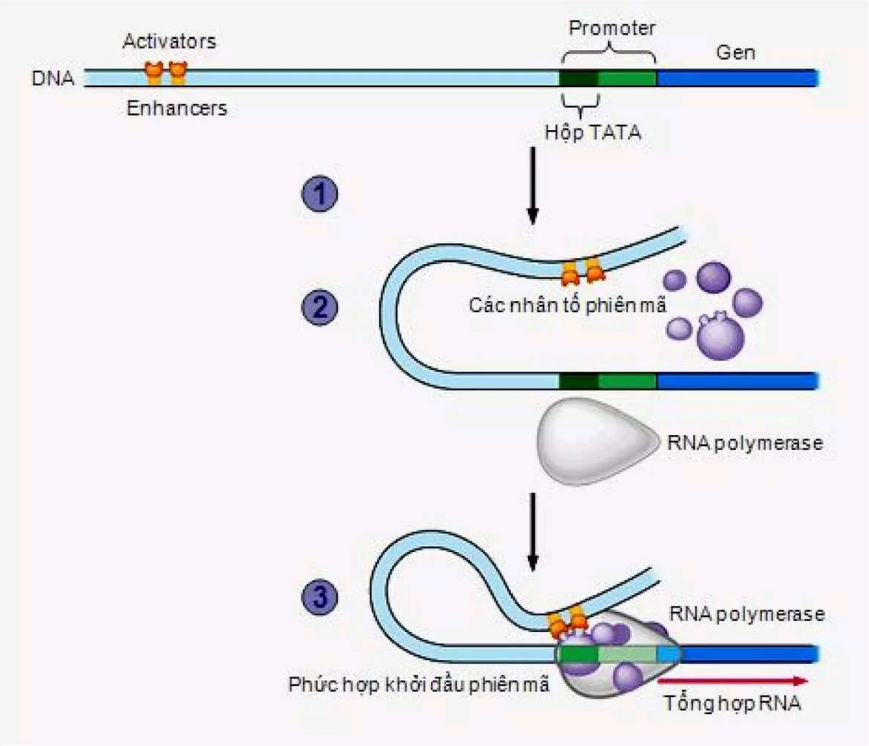
3. Các protein là nhân tố có tác động trans

Một vài protein, nhận biết hộp CCAAT, đã được xác định ở tế bào động vật có vú. Các nhân tố này có thể được phân biệt giữa các đơn vị của các phân tử CCAAT. Các hộp GC được nhận biết bởi các nhân tố phụ. Các ví dụ về các tác động *trans* (*trans*-acting factor) đã tìm thấy ở nấm men và động vật có vú. Các nhân tố *trans* có đặc điểm chung là gồm hai vùng cấu trúc và chức năng chính:

- Vùng gắn nhân tố *trans* vào DNA.
- Vùng tác động lên sự phiên mã.

Đồng thời các nhân tố *trans* có bốn kiểu tác động như:

- "ngón tay kẽm" (zinc-finger)
- "xoắn-vòng-xoắn" (helix-turn-helix)
- "xoắn-nút-xoắn" (helix-loop-helix)
- "dây kéo leucine" (leucine-zipper) (Hình 8.9).



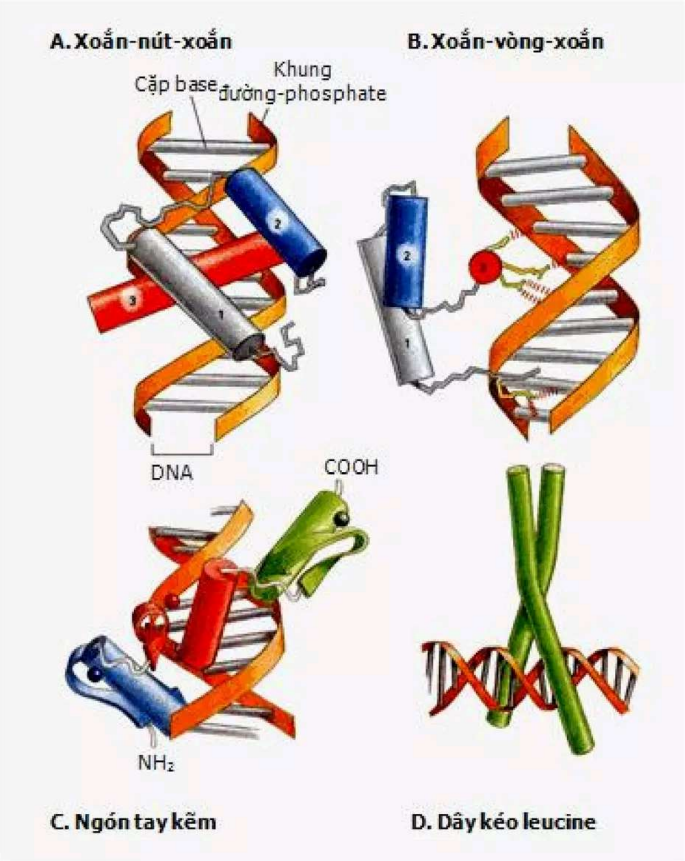
Hình 8.8. Enhancer của DNA eukaryote

Đặc tính chung của các cấu trúc vừa kể là chúng luôn luôn gắn lên các trình tự *cis* dưới dạng dimer (gồm hai dimer).

Trình tự *cis* và nhân tố *trans* thường có các tương tác đặc hiệu do những liên kết yếu hình thành giữa các phân tử của nhân tố *trans* với các base của trình tự *cis*. Ví dụ, trường hợp tương tác giữa các hormone steroid với các enhancer.

Nhiều protein có tác động *trans* tham gia vào tiến trình phát triển của cơ thể ruồi giấm, chúng có vai trò đặc biệt trong xác định sự sắp xếp các đoạn thân của ruồi.

Như vậy, các gen eukaryote được hoạt hóa bởi hai trình tự DNA có tác động *cis* là promoter và enhancer, chúng nhận biết được các nhân tố protein có tác động *trans*, các nhân tố *trans* cho phép RNA polymerase khởi sự phiên mã và đạt tốc độ phiên mã tối đa.



Hình 8.9. Các kiểu tác động của nhân tố *trans*

4. Hormone

Ví dụ rõ nhất về các chất điều hòa nội tại của hoạt tính gen là hormone. Đó là những chất tạo ra do một loạt tế bào có hiệu quả đến các tế bào khác. Các hormone thường được vận chuyển đến các phần của cơ thể nhưng chỉ tác động đến các tế bào có các receptor tương ứng.

Sự tương tác giữa hormone với receptor gây ra tín hiệu tác động đến các vùng đặc trưng của DNA, gây hoạt hóa gen hoặc một nhóm gen tương ứng.

Các hormone có thể kích thích phiên mã bởi một trong các cơ chế sau:

- Hormone có thể làm cho DNA tách khỏi histone và tạo điều kiện cho RNA polymerase bắt đầu phiên mã.

- Có thể làm chất cảm ứng (inducer) gây bất hoạt phân tử receptor.

- Có thể gắn trực tiếp với đoạn DNA đặc hiệu tạo thuận lợi cho việc gắn RNA polymerase hoặc protein là nhân tố phiên mã (protein transcription factor).

- Có thể hoạt hóa effector protein làm thành phức hợp gắn DNA và kích thích gắn RNA polymerase.

- Có thể gắn với protein tạo phức hợp hoạt hóa kích thích gắn RNA polymerase.

5. Kiểm soát các chất thường gặp trong nhân

Một số nhóm phân tử hiện diện rất nhiều trong mỗi tế bào eukaryote, như các histone, các chất của bộ máy dịch mã, các cấu phần của màng… Vài chương trình khác nhau được thực hiện để duy trì số lượng đủ lớn của chúng trong tế bào. Các chương trình đó gồm:

- Sự phiên mã liên tục và lặp lại trong tế bào.

- Sự lặp lại các gen.

- Sự khuếch đại ngoài nhiễm sắc thể các trình tự gen đặc hiệu.

5.1. Sự phong phú

Các gen mã hóa cho rRNA và tRNA đã được nghiên cứu kỹ. Phản ứng lai đơn giản DNA-RNA cho thấy nhiễm sắc thể của *Xenopus* có vùng tổ chức hạch nhân (nucleolus organizer) chứa 450 bản sao của DNA mã hóa cho 18S và 28S rRNA. Ngược lại, trong mỗi nhân có 20.000 bản sao của các gen mã hóa cho 5S rRNA và các gen này không nằm ở vùng tổ chức hạch nhân.

Các gen mã hóa cho histone cũng có với nhiều bản sao được lặp lại hàng trăm lần.

Sự phong phú của các gen nêu trên đảm bảo đủ số lượng cần thiết cho dịch mã khi phải tổng hợp hàng triệu phân tử protein vào lúc cần thiết.

5.2. Sự khuếch đại gen

Một ví dụ về sự khuếch đại gen trong nhân là các chỗ phình (puff) của nhiễm sắc thể khổng lồ hay đa sợi (polytene chromosome) ở tuyến nước bọt của ruồi giấm. Ở các chỗ phình này, DNA nguyên nhiễm sắc (euchromatic) được khuếch đại khoảng vài nghìn lần.

Trong nhân tế bào trứng của *Xenopus*, có hàng trăm nhân con ngoài nhiễm sắc thể (extrachromosome nucleoli) với kích thước khác nhau, mỗi nhân con chứa các vòng tròn rDNA kích thước khác nhau, mà vai trò đến nay chưa rõ. Các DNA vòng tròn này sản sinh nhiều rRNA ở mức hợp lý.

V. Sự biệt hóa tế bào

1. Các tế bào biệt hóa mang thông tin giống nhau

Ở các sinh vật bậc cao cũng như ở người, cơ thể trưởng thành gồm nhiều loại tế bào khác nhau. Các tế bào này đều bắt nguồn từ một hợp tử ban đầu, nhưng đã qua quá trình biệt hóa trở thành các tế bào có các chức năng khác nhau. Tuy nhiên, thực nghiệm xác định rằng số lượng nhiễm sắc thể, số lượng DNA và cả tỷ lệ (A+T)/(G+C) của các tế bào thuộc các mô khác nhau của cùng một cơ thể đều giống nhau. Sử dụng kỹ thuật lai DNA cho thấy DNA từ những tế bào của các mô khác nhau của cùng một cơ thể không bị biến đổi trong quá trình biệt hóa, chúng có thể tự hồi tính (renaturation) với nhau.

Thí nghiệm ghép nhân tế bào ruột ếch vào tế bào trứng bị hỏng nhân (do chiếu tia tử ngoại) cho thấy 1% tế bào ghép nhân phát triển thành ếch trưởng thành. Điều này chứng tỏ tế bào ruột tuy đã biệt hóa vẫn giữ nguyên vẹn thông tin di truyền để tạo ra ếch trưởng thành.

Tóm lại, số lượng DNA của các tế bào biệt hóa về căn bản giống với các hợp tử ban đầu và chứa nguyên vẹn thông tin di truyền đủ để phát triển thành cá thể nguyên vẹn.

2. Các tế bào biệt hóa tổng hợp các nhóm protein khác nhau

Để hiểu được sự khác nhau giữa các tế bào biệt hóa, cần xét nhiều vấn đề sau:

- Thứ nhất, nhiều quá trình là chung cho tất cả các tế bào nên có nhiều protein giống nhau. Những protein này gồm số lượng nhiều, dễ phân tích như phần lớn các protein cấu trúc của thành tế bào và nhiễm sắc thể; một số protein căn bản của các bào quan (lưới nội sinh chất, bộ máy Golgi, các ribosome…). Nhiều protein có số lượng không nhiều như các enzyme khác nhau, tham gia vào các phản ứng trung tâm của quá trình trao đổi chất, đều hiện diện như nhau ở tất cả các kiểu tế bào.

- Thứ hai, một số protein có số lượng phong phú ở một số tế bào chuyên hóa mà việc phát hiện chúng cần có thử nghiệm riêng. Ví dụ, hemoglobin chỉ có thể phát hiện ở tế bào máu.

- Thứ ba, nếu như hơn 2.000 loại protein có số lượng dồi dào được so sánh giữa các kiểu tế bào biệt hóa của cùng một vi sinh vật với nhau bằng điện di hai chiều polyacrylamide, thì một số khác biệt đáng kể sẽ được tìm thấy. Dù sự so sánh giữa hai dòng tế bào nuôi cấy (như các dòng tế bào cơ và thần kinh) hay giữa các tế bào của cùng một mô của chuột (như gan và phổi), đại đa số các protein được phát hiện đều được tổng hợp ở cả hai loại tế bào và với tốc độ tổng hợp có thể khác nhau đến 10^5; chỉ có vài trăm protein hiện diện với số lượng khác nhau đáng kể ở hai kiểu tế bào.

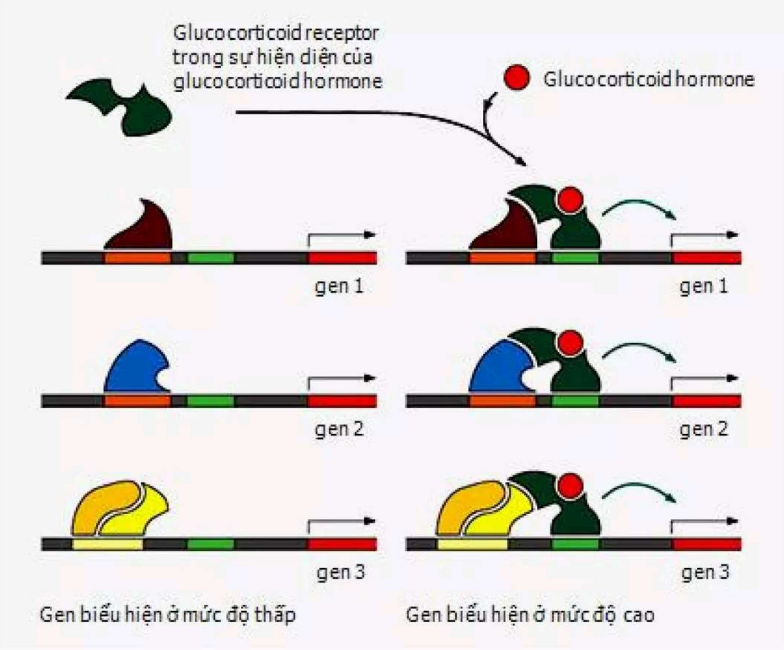
Các nghiên cứu cho thấy các tế bào eukaryote bậc cao tổng hợp từ 10.000-20.000 protein khác nhau. Phần lớn chúng có số lượng rất ít và khó phát hiện. Nếu các protein hiếm (minor protein) này khác nhau giữa các tế bào với mức độ tương tự các protein dồi

dào, thì chỉ một số lượng nhỏ của chúng cũng đủ tạo nên nhiều khác biệt chuyên hóa lớn về hình thái và sinh lý của tế bào.

Tế bào có thể thay đổi sự biểu hiện gen của chúng đáp lại tín hiệu bên ngoài.

Phần lớn các tế bào chuyên hóa của sinh vật đa bào có khả năng thay đổi phương thức biểu hiện của gen để đáp lại tác động bên ngoài. Ví dụ: nếu tế bào bị tác động bởi glucocorticoid hormone, thì sự tổng hợp một số protein chuyên biệt được tăng vọt (Hình 8.10). Glucocorticoid được phóng thích ra khi bị đói bụng hay hoạt động mạnh và báo hiệu cho gan tăng cường tạo glucose từ amino acid hay các phân tử nhỏ khác nhau. Các protein, được tạo ra do các enzyme cảm ứng như tyrosine aminotransferase hỗ trợ biến tyrosine thành glucose. Khi hormone không còn nữa, sự tổng hợp các protein này giảm xuống mức bình thường.

Các loại tế bào khác nhau phản ứng với glucocorticoid bằng nhiều cách khác nhau. Ví dụ ở tế bào mỡ, sự tổng hợp tyrosine transferase giảm, trong khi đó một số loại tế bào khác hoàn toàn không có phản ứng với glucocorticoid. Các ví dụ này minh họa cho tính chất chung của biệt hóa là các kiểu tế bào biệt hóa khác nhau thường phản ứng lại cùng một tín hiệu bên ngoài bằng nhiều cách khác nhau. Trên cơ sở đó, mỗi kiểu tế bào biệt hóa có một đặc tính ổn định thường xuyên. Các tính chất đó phản ánh sự biểu hiện lâu bền của các nhóm gen khác nhau.



Hình 8.10. Tác dụng của glucocortcoid làm tăng mức độ biểu hiện của gen

Như vậy, các tế bào biệt hóa chỉ sử dụng một phần thông tin. Nhiều loại tế bào chuyên hóa tổng hợp chủ yếu một số protein, ngoài các protein cấu trúc và protein cơ bản sử dụng cho các quá trình sinh lý bình thường. Ví dụ: tế bào cơ tổng hợp nhiều myosin, một protein quan trọng trong co cơ, hay tế bào biểu bì (epithelial) tổng hợp nhiều keratine. Như vậy, cũng như chứa thông tin di truyền giống nhau, nhưng mỗi loại tế bào biệt hóa chỉ sử dụng một phần thông tin: tổng hợp chủ yếu một số loại protein .

3. Sự điều hòa ở mức phiên mã là nguồn gốc căn bản của các sai khác giữa những tế bào biệt hóa

Nếu những sự biệt hóa giữa các kiểu tế bào khác nhau phụ thuộc vào các gen chuyên biệt mà tế bào biểu hiện, thì sự kiểm soát biểu hiện gen được thực hiện ở mức độ nào?

Giả thuyết được chấp nhận hiện nay là trong các tế bào biệt hóa một số gen phiên mã, còn có các gen khác thì không. Không có sự kiện nào mâu thuẫn với giả thuyết này và nó giải thích hợp lý hơn cả tình trạng biệt hóa của các tế bào. Đối với phần lớn gen, sự kiểm soát phiên mã có tầm quan trọng hàng đầu, vì chỉ nó mới đảm bảo cho sự tổng hợp không dư thừa các chất trung gian.

Việc phát hiện các gen điều hòa và các gen đóng hay mở giúp hiểu được sự điều hòa quá trình phát triển cá thể và biệt hóa tế bào. Genome đơn bội của tế bào người có số lượng DNA nhiều hơn gấp 1.000 lần so với genome của vi khuẩn. Tuy nhiên, số lượng gen cấu trúc ở người chỉ lớn hơn 10 lần số gen cấu trúc vi khuẩn. Điều đó cho thấy nhiều gen ở người tham gia vào các cơ chế điều hòa.

Tóm lại, genome của tế bào chứa các trình tự nucleotide của DNA thông tin để tạo ra hàng nghìn các protein và phân tử RNA khác nhau. Tế bào chỉ biểu hiện điển hình một nhóm gen của nó và các kiểu tế bào biệt hóa khác nhau ở sinh vật đa bào được sản sinh ra do các nhóm gen khác nhau có sự biểu hiện không giống nhau. Hơn thế nữa, các tế bào có thể thay đổi phương thức biểu hiện các gen của chúng để đáp lại những thay đổi của môi trường, các tín hiệu đó từ các tế bào khác. Mặc dù tất cả các bước trong sự biểu hiện của gen về căn bản đều được điều hòa, nhưng đối với phần lớn các gen việc khởi sự phiên mã là điểm kiểm soát quan trọng nhất.

Tài liệu tham khảo/đọc thêm

1. Phạm Thành Hổ. 2003. Di truyền học. *NXB Giáo dục,* Hà Nội.

2. Lê Đức Trình. 2001. Sinh học phân tử của tế bào. *NXB Khoa học và Kỹ thuật*, Hà Nội.

3. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

4. Karp G. 2002. Cell and Molecular Biology: Concepts and Experiments. 3rd ed. *John Wiley & Sons, Inc*. New York, USA.

5. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

6. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *Freeman and Company*, New York, USA.

7. Walker JM and Rapley R. 2000. Molecular Biology and Biotechnology. *Chapman & Hall Limited*, London, UK.

8. Watson JD, Hopkins NH, Roberts JW and Weiner AM. 2004. Molecular Biology of the Gene. *The Benjamin/Cummings Publishing Company, Inc*. California, USA.

9. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company*, New York, USA

Chương 9

Công nghệ DNA tái tổ hợp

I. Mở đầu

Vào năm 1973, một nhóm các nhà khoa học đã tạo ra cơ thể sinh vật đầu tiên với các phân tử DNA tái tổ hợp. Theo đó, Cohen (ĐH Stanford, Mỹ) và Boyer (ĐH California, Mỹ) cùng các cộng sự đã đưa được một đoạn DNA từ một plasmid này vào một plasmid khác, tạo ra một plasmid hoàn toàn mới, plasmid tái tổ hợp. Sau đó, họ đưa plasmid tái tổ hợp vào trong các tế bào *E. coli*. Trong một thời gian ngắn, các tác giả này đã dùng các phương pháp giống nhau để gắn các gen từ hai loại vi khuẩn khác nhau, cũng như để chuyển các gen từ ếch vào vi khuẩn. Các thí nghiệm này đánh dấu một cuộc cách mạng vô cùng quan trọng trong lịch sử nghiên cứu khoa học của nhân loại.

Công nghệ DNA tái tổ hợp là một tập hợp các kỹ thuật phân tử để định vị, phân lập, biến đổi và nghiên cứu các đoạn DNA. Thuật ngữ tái tổ hợp được dùng thường xuyên do mục tiêu của nó là phối hợp DNA từ hai nguồn xa nhau. Ví dụ: các gen từ hai nguồn vi khuẩn khác nhau có thể được liên kết lại, hoặc một gen người có thể được đưa vào nhiễm sắc thể vi khuẩn. Công nghệ DNA tái tổ hợp (còn gọi là công nghệ di truyền, công nghệ gen hay kỹ thuật gen…) hiện nay bao gồm một mạng lưới các kỹ thuật phân tử được dùng để phân tích, biến đổi và tái tổ hợp hầu như mọi trình tự DNA.

1. Tác động của công nghệ DNA tái tổ hợp

Công nghệ DNA tái tổ hợp đã biến đổi sâu sắc phương thức nghiên cứu gen. Trước đây, thông tin về cấu trúc và tổ chức của gen thu được bằng cách kiểm tra biểu hiện kiểu hình của chúng, nhưng những kỹ thuật mới đã tạo ra khả năng tự đọc các trình tự nucleotide. Trước đây, các nhà di truyền phải chờ đợi sự xuất hiện các đột biến ngẫu nhiên hoặc cảm ứng để phân tích hiệu quả của sự sai khác di truyền, ngày nay họ có thể tạo ra đột biến ở các điểm nhất định một cách chính xác và xem chúng thay đổi kiểu hình như thế nào.

Công nghệ DNA tái tổ hợp đã cung cấp các thông tin mới về cấu trúc và chức năng của gen và đã thay đổi nhiều khái niệm cơ bản của di truyền học. Ví dụ: trong khi mã di truyền được xem là rất phổ biến, thì bây giờ chúng ta còn biết rằng các mã không phổ biến cũng tồn tại trong DNA ty thể. Trước đây, chúng ta nghĩ rằng tổ chức của các gen eukaryote giống với prokaryote, nhưng bây giờ chúng ta biết rằng nhiều gen eukaryote bị gián đoạn bởi các intron. Ngày nay, chúng ta đã biết đầy đủ hơn về các quá trình tái bản, phiên mã, dịch mã, biến đổi RNA (RNA processing) và điều hòa gen thông qua việc sử dụng các kỹ thuật tái tổ hợp DNA. Các kỹ thuật này cũng được dùng trong nhiều trong nhiều lĩnh vực khác, bao gồm hóa sinh học, vi sinh vật học, sinh học phát triển, sinh học thần kinh, tiến hóa và sinh thái học.

Công nghệ DNA tái tổ hợp cũng được ứng dụng để tạo ra nhiều sản phẩm thương mại, chẳng hạn: thuốc, hormone, enzyme và các giống cây trồng-vật nuôi. Một nền công nghiệp hoàn toàn mới, công nghiệp công nghệ sinh học, đã phát triển chung quanh việc sử dụng các kỹ thuật này để tạo ra các sản phẩm mới. Trong y học, các kỹ thuật tái tổ hợp DNA được dùng để thăm dò bản chất của ung thư, chẩn đoán các bệnh di truyền và nhiễm trùng, sản xuất thuốc và điều trị các rối loạn di truyền.

2. Làm việc ở mức độ phân tử

Kỹ thuật gen cho thấy một loạt cơ hội, mở ra các phương thức cần thiết (mà trước đây có thể không được) gần như là hiển nhiên. Vấn đề cơ bản đó là các gen có kích thước quá nhỏ và có hàng ngàn gen ở trong mỗi tế bào. Thậm chí, khi quan sát trên kính hiển vi mạnh nhất, thì DNA xuất hiện như là một sợi dây bé xíu, các nucleotide riêng rẽ không thể thấy, và không có một dấu hiệu nào về các đường nét vật lý ở chỗ bắt đầu và kết thúc của một gen.

Để minh họa vấn đề này, chúng ta hãy xem xét một ví dụ đặc trưng về di truyền phân tử như sau: Giả thiết rằng chúng ta muốn phân lập một gen đặc biệt của người và đặt nó vào trong vi khuẩn để sản xuất một lượng lớn các protein người đã được mã hóa. Vấn đề đầu tiên là tìm được gen mong muốn. Genome đơn bội của người chứa khoảng 3,3 tỷ cặp base của DNA. Giả sử gen mà chúng ta muốn phân lập dài 3.000 bp. Như vậy, gen đích của chúng ta chỉ chiếm một phần triệu của genome; vì thế để tìm kiếm gen của chúng ta trong một genome đồ sộ như thế là khó khăn hơn rất nhiều so với việc tìm kiếm

một cây kim trong một đống cỏ khô. Nhưng thậm chí, nếu chúng ta có thể định vị gen, thì chúng ta sẽ tách nó ra khỏi genome như thế nào? Không có forcept đủ nhỏ để gắp một mảnh DNA đơn, và cũng không có một cái kéo cơ học nào đủ nhỏ để cắt ra khỏi genome một đoạn gen riêng biệt.

Nếu chúng ta thành công trong việc định vị và phân lập gen mong muốn, thì bước tiếp theo chúng ta cần đưa nó vào trong tế bào vi khuẩn. Các đoạn DNA mạch thẳng sẽ bị thoái biến nhanh bởi vi khuẩn; vì thế gen phải được chèn vào trong một dạng ổn định. Nó cũng phải ổn định để tái bản thành công hoặc nó sẽ không được phân chia tiếp khi tế bào phân chia.

Nếu chúng ta chuyển gen vào vi khuẩn thành công trong một dạng ổn định, chúng ta vẫn còn phải đảm bảo rằng gen được phiên mã và dịch mã. Sự biểu hiện của gen là một quá trình phức tạp đòi hỏi một số các trình tự DNA khác nằm ở bên ngoài gen. Tất cả những trình tự này phải hiện diện trong các hướng ở các vị trí thích hợp của chúng để sản xuất protein.

Cuối cùng, các phương pháp được sử dụng để phân lập và chuyển gen có hiệu quả vô cùng thấp, trong hàng triệu tế bào được hướng tới cho các phương thức này, chỉ có một tế bào có thể chọn lọc thành công và biểu hiện gen của người. Vì thế, chúng ta phải tìm kiếm nhiều tế bào vi khuẩn để phát hiện được một tế bào mang DNA tái tổ hợp.

Trước đây, các vấn đề này dường như là không vượt qua được. Nhưng ngày nay, các kỹ thuật phân tử được phát triển để khắc phục chúng, và các gen người được chuyển dễ dàng vào các tế bào vi khuẩn và ở đó chúng sẽ được biểu hiện tốt.

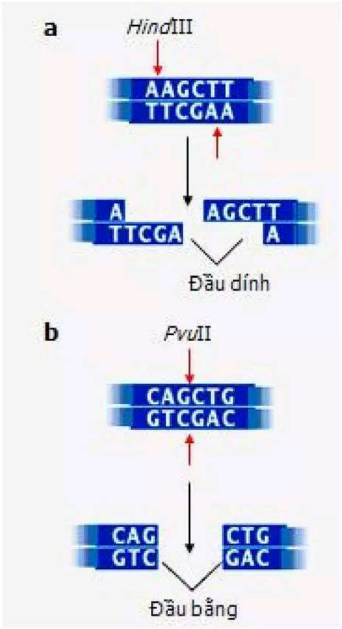
II. Endonuclease hạn chế

Trong tự nhiên, các enzyme endonuclease hạn chế (restriction endonuclease, RE), gọi tắt là enzyme hạn chế, hiện diện trong hầu hết các tế bào vi khuẩn để ngăn cản DNA ngoại lai tiếp quản bộ máy tổng hợp protein của tế bào. DNA của chính chúng sẽ được bảo vệ khỏi tác dụng của enzyme hạn chế nhờ sự có mặt của các enzyme nội bào có thể methyl hóa (methylation) các nucleotide đặc biệt, vì thế các nucleotide này không được nhận biết bởi các enzyme hạn chế.

Việc phát hiện ra các enzyme hạn chế của vi khuẩn cắt DNA ở những trình tự đặc biệt, đã giúp cho việc thao tác gen dễ dàng hơn, do nó có thể giảm chiều dài của các phân tử DNA thành một tập hợp bao gồm các đoạn ngắn hơn. Hiện nay, các enzyme hạn chế được phân lập từ các sinh vật prokaryote.

Mỗi enzyme hạn chế chỉ nhận biết và cắt một trình tự DNA đặc biệt thường chứa bốn hoặc sáu cặp nucleotide. Ví dụ enzyme *Eco*RI tách chiết từ *E. coli* cắt trình tự GAATTC, enzyme *Bal*I của *Brevibacterium albidum* cắt trình tự TGGCCA. Có hơn 900

enzyme hạn chế khác nhau được tinh sạch từ khoảng 250 chủng vi sinh vật. Các enzyme hạn chế cắt các phân tử DNA sợi đôi theo hai cách khác nhau (Hình 9.1):



Hình 9.1. Hai kiểu cắt của enzyme hạn chế. (a) tạo ra đầu so le, và (b) tạo ra đầu bằng.

- Cắt trên một đường thẳng đối xứng để tạo ra các phân tử đầu bằng (đầu thô).

- Cắt trên những vị trí nằm đối xứng quanh một đường thẳng đối xứng để tạo ra những phân tử đầu so le (đầu dính).

Vì một enzyme hạn chế chỉ nhận biết một trình tự duy nhất, cho nên số vị trí cắt trên một phân tử DNA đặc biệt thường là nhỏ. Các đoạn DNA được cắt bởi enzyme hạn chế có thể được phân tách theo kích thước bằng điện di agarose gel để nghiên cứu. Do sự tương tự của tổ chức phân tử trong tất cả các cơ thể, cho nên DNA vi khuẩn, DNA thực vật và DNA động vật có vú tương hợp nhau về cấu trúc. Vì thế, một đoạn DNA từ một dạng sống này có thể dễ dàng được pha trộn với DNA của một dạng sống khác. Sự tương tự này cũng phù hợp đối với plasmid, nhân tố di truyền ngoài nhân được tìm thấy trong nhiều loài vi khuẩn khác nhau. Chúng là những phân tử DNA mạch vòng đóng sợi đôi được dùng làm vector mang các đoạn DNA ngoại lai dùng trong kỹ thuật tái tổ hợp DNA.

Giống như *Eco*RI, nhiều enzyme hạn chế đã tạo ra các đoạn DNA với đầu 5’ lồi (protruding). Một số enzyme hạn chế (ví dụ: *Pst*I) đã tạo ra các đoạn DNA có đầu 3’ lồi. Một số enzyme hạn chế khác (ví dụ: *Bal*I) cắt ở trục đối xứng để tạo ra các đoạn DNA mang đầu bằng (blunt) (Bảng 9.1).

Bảng 9.1. Trình tự nhận biết của một số enzyme hạn chế

Enzyme	Nguồn vi sinh vật	Trình tự nhận biết	Loại đầu
*Bam*HI	*Bacillus amyloliquefaciens*	5'-G GATCC-3' 3'-CCTAG G-5'	Dính
*Bgl*II	*Bacillus globigii*	5'-A GATCT-3' 3'-TCTAG A-5'	Dính
*Cof*I	*Clostridium formicoaceticum*	5'-G CGC-3' 3'-CGC G-5'	Dính
*Dra*I	*Deinococcus radiophilus*	5'-TTT AAA-3' 3'-AAA TTT-5'	Bằng
*Eco*RI	*Escherichia coli*	5'-G AATTC-3' 3'-CTTAA C-5'	Dính
*Hae*III	*Haemophilus aegypticus*	5-GG CC-3' 3'-CC GG-5'	Bằng
*Hind*III	*Haemophilus influenzae*	5-A AGCTT-3' 3'-TTCGA A-5'	Dính
*Hpa*II	*Haemophilus parainfluenzae*	5'-C CGG-3' 3'-GGC C-5'	Dính
*Pst*I	*Providencia stuartii*	5'-CTGCA G-3' 3'-G ACGTC-5'	Dính
*Pvu*II	*Protrus vulgaris*	5'-CAG CTG-3' 3'-GTC GAC-5'	Bằng
*Sma*I	*Serratia marcescens*	5'-CCC GGG-3' 3'-GGG CCC-5'	Bằng
*Xma*I	*Xanthomonas malvacearum*	5'-C CCGGG-3' 3'-GGGCC C-5'	Dính

1. Gắn các đầu bị cắt bởi enzyme hạn chế

Có hai kiểu gắn khác nhau: gắn đầu bằng và gắn đầu dính, bằng cách dùng enzyme DNA ligase của bacteriophage T4 có thể gắn các đầu bằng hoặc đầu dính lại với nhau, tuy nhiên trường hợp đầu bằng thường cho hiệu quả thấp hơn đầu dính.

Ví dụ: Hình 9.2 minh họa việc gắn các đầu dính được cắt bằng enzyme *Hind*III.

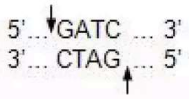
2. Isochizomer

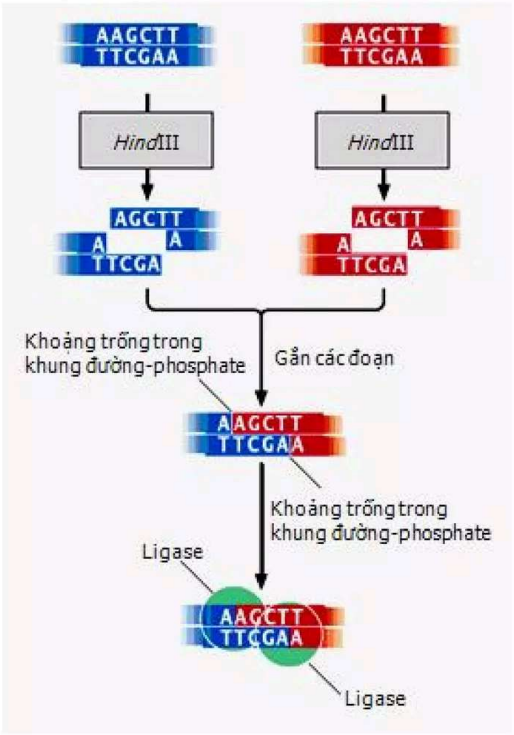
Nói chung, các RE khác nhau nhận biết các trình tự khác nhau (Bảng 9.1), tuy nhiên cũng có một số trường hợp cho thấy có những enzyme được phân lập từ nhiều nguồn khác nhau nhưng lại cắt trong một trình tự, các enzyme đó được gọi là

isochizomer. Hơn nữa, một vài enzyme nhận biết các chuỗi tetranucleotide mà trong một số trường hợp các tetranucleotide này lại xuất hiện trong các chuỗi hexanucleotide của các enzyme khác.

Ví dụ:

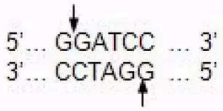
- *Mbo*I và *Sau*3A nhận biết trình tự:





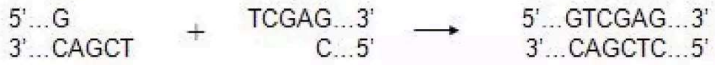
Hình 9.2. Gắn các đầu dính

- Trong khi đó *Bam*HI nhận biết trình tự:



Trong một vài trường hợp các đoạn được tạo ra nhờ một enzyme hạn chế này khi gắn với các đoạn được tạo ra nhờ một enzyme hạn chế khác đã hình thành các thể lai mà các enzyme bố mẹ không thể nhận biết được. Ví dụ: *Sal*I cắt trình tự (G↓TCGAC) và *Xho*I cắt trình tự (C↓TCGAG), các trình tự được sinh ra này đã gắn với nhau tạo thành

thể lai (hybrid) mà các vị trí cắt hạn chế của chúng không thể nhận biết bởi các enzyme *Sal*I và *Xho*I:



III. Phương thức tạo dòng

Các phương thức cơ bản của kỹ thuật DNA tái tổ hợp là: (1) Gắn một đoạn DNA vào một phân tử DNA (như là vector) có thể tái bản, và (2) cung cấp một môi trường cho phép sao chép phân tử DNA đã được gắn.

Có ba nhóm vector được dùng phổ biến để tạo dòng các đoạn DNA ngoại lai và tái bản (sao chép) trong *E. coli*; đó là plasmid, bacteriophage λ và cosmid. Tất cả những vector này phải có một số tính chất cần thiết sau:

- Chúng có khả năng tự tái bản trong *E. coli*.

- Mang các gen chỉ thị chọn lọc để dễ dàng phân biệt và tinh sạch vector của thể tái tổ hợp với các dạng khác.

- Chúng có các vùng DNA không cần thiết cho sự sinh sản trong vi khuẩn, vì thế DNA ngoại lai có thể được đưa vào trong các vùng này.

- Chúng có thể biến nạp vào tế bào vật chủ một cách dễ dàng.

1. Plasmid vector

Plasmid là các thông tin di truyền ngoài nhân có trong nhiều loài vi khuẩn khác nhau. Chúng là những phân tử DNA mạch vòng, sợi đôi, kích thước từ 1- ≥ 200 kb, có khả năng tự sao chép để tồn tại độc lập trong tế bào. Các plasmid có thể mang một số gen của vi khuẩn và các gen này có thể biểu hiện ra protein.

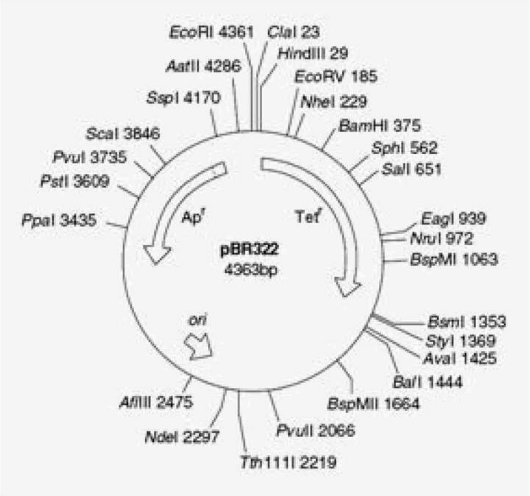
DNA của plasmid có thể được phân lập từ nuôi cấy vi khuẩn chứa plasmid bằng cách bổ sung chất tẩy (như là sodium dodecyl sulfate-SDS) và ly tâm sự sinh tan (lysate)[1]. Phức hợp nhiễm sắc thể vi khuẩn, lớn hơn plasmid nhiều, sẽ lắng xuống đáy của tube ly tâm, plasmid siêu xoắn và các đoạn nhiễm sắc thể mạch thẳng giữ lại trong thể nổi. Plasmid siêu xoắn một lần nữa được phân tách bằng ly tâm sau khi xử lý với CsCl và EtBr.

Plasmid mang các gen mã hóa cho các enzyme thường có lợi cho vi khuẩn vật chủ. Các plasmid có thể mang các kiểu hình khác nhau như: kháng kháng sinh, sản xuất kháng sinh, phân hủy các hợp chất hữu cơ phức tạp, sản xuất các enzyme hạn chế và enzyme biến đổi (modification enzymes).

Các plasmid có thể được chuyển vào trong vi khuẩn sau khi vi khuẩn được xử lý để tế bào có thể cho thấm qua nhất thời đối với các phân tử DNA nhỏ. Quá trình này được biết như là sự biến nạp (transformation). Vi khuẩn được biến nạp thành công có thể được chọn lọc dựa trên kiểu hình mới mà chúng nhận được từ plasmid, chẳng hạn khả năng kháng các kháng sinh.

Một số plasmid hiện diện trong tế bào có số bản sao thấp, một hoặc một vài bản sao trên tế bào, do DNA của plasmid chỉ sao chép một hoặc hai lần trước khi tế bào phân chia. Tuy nhiên, các plasmid khác tồn tại một số bản sao lớn hơn (10 tới 100 bản sao trên một tế bào) do DNA tái bản lặp lại cho đến khi đạt được số bản sao thích hợp. Các plasmid có số bản sao lớn được gọi là plasmid dạng xoắn lỏng lẻo (relaxed plasmid), và đây là một trong những tính chất hữu ích của vector tạo dòng.

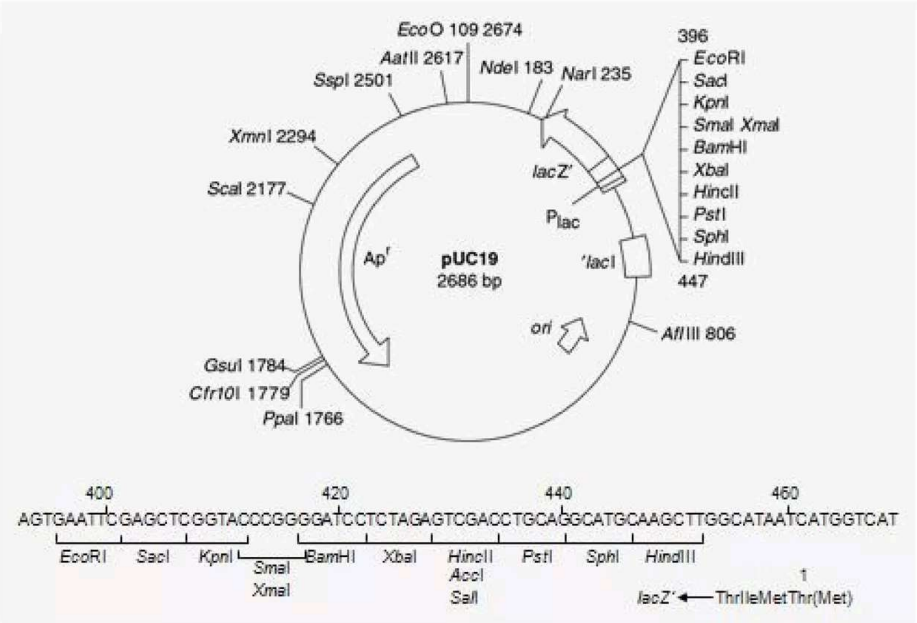
Hình 9.3 trình bày một trong các plasmid vector thế hệ thứ hai dạng xoắn lỏng lẻo, pBR322, dài 4.363 bp, vector này chứa hai gen kháng kháng sinh là ampicillin (Amp) và tetracycline (Tet). Số thứ tự của các nucleotide trên vector được bắt đầu với vị trí *Eco*RI đơn: T đầu tiên trong chuỗi GAATTC được quy ước là nucleotide thứ nhất. Các số thứ tự sau đó được tiếp tục quanh phân tử vector theo hướng từ gen kháng tetracycline tới gen kháng ampicillin.



Hình 9.3. Plasmid vector pBR322. Ap^r (hay Amp^r) và Tet^r: gen kháng ampicillin và tetracycline, *ori*: trình tự khởi đầu sao chép, và một số vị trí nhận biết cho các RE.

Hình 9.4. trình bày một loại plasmid vector thế hệ thứ ba là pUC19, đây là loại vector tạo dòng đặc trưng, Nó mang vùng tạo dòng (multiple cloning sites) hay còn gọi là vùng đa nối (polylinker), vùng khởi đầu sao chép (*ori*), và hai gen chỉ thị (gen kháng ampicillin và gen *lacZ'*). Ampicillin là loại kháng sinh giết chết tế bào vi khuẩn, nhưng những vi khuẩn nào chứa vector pUC19 sẽ kháng lại loại kháng sinh này. Gen *lacZ'* mã

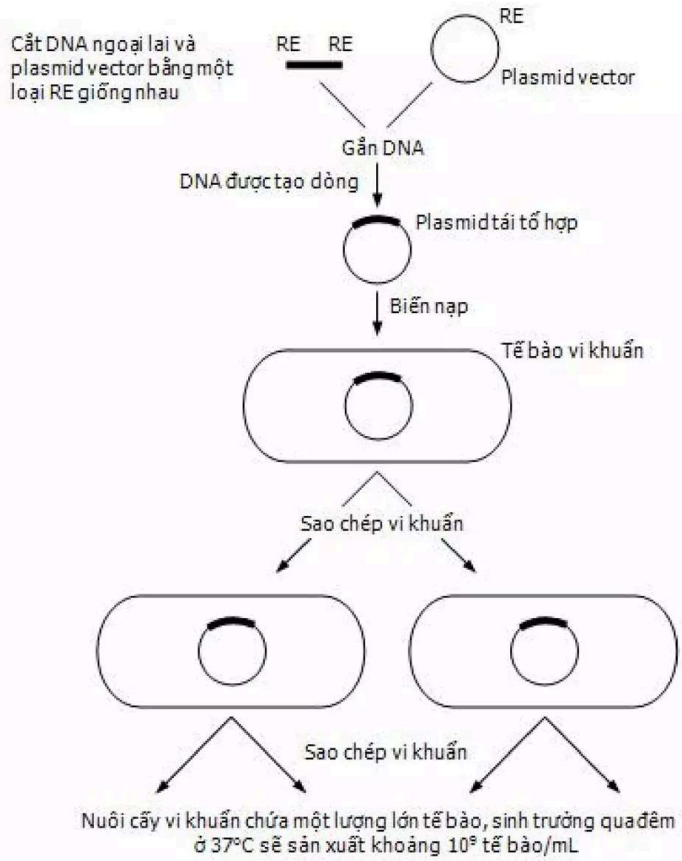
hóa enzyme β-galactosidase, bình thường enzyme này cắt lactose để sản xuất ra glucose và galactose. Enzyme này cũng cắt X-gal để tạo ra một cơ chất màu xanh; khi X-gal được bổ sung vào môi trường, các khuẩn lạc vi khuẩn chứa pUC19 sẽ có màu xanh và dễ dàng nhận biết. Vùng polylinker của vector pUC19 là tập hợp một số vị trí nhận biết đơn của các enzyme hạn chế cho phép gắn đoạn DNA ngoại lai vào plasmid.



Hình 9.4. Plasmid vector tạo dòng đặc trưng pUC19. Mang các vị trí cắt hạn chế đơn trong vùng tạo dòng, vùng khởi đầu sao chép (*ori*), và hai gen chỉ thị (gen Ap^r và gen *lacZ'*).

Plasmid có thể được cắt ở một vị trí xác định bằng enzyme hạn chế. Vì thế, các đoạn được tạo ra có thể tạo vòng bằng cách kết hợp các đầu dính bổ sung. Hơn nữa, các đoạn được tạo ra bởi một enzyme đặc biệt hoạt động trên một phân tử DNA sẽ có đầu tương đồng (đầu dính) với đoạn được tạo ra bởi cùng một enzyme hoạt động trên một phân tử DNA khác. Vì thế, các đoạn từ hai phân tử DNA khác nhau, từ hai cơ thể khác nhau có thể kết hợp bằng các liên kết hydrogen thuận nghịch khi các đoạn này được trộn lại.

Nếu sự kết hợp được gắn lại sau khi bắt cặp, thì các đoạn được kết hợp cố định bền vững, sự kết hợp của các đoạn này được thực hiện nhờ enzyme DNA ligase (còn gọi là polynucleotide ligase) liên kết cộng hóa trị các phân tử tái tổ hợp bằng cách tạo ra liên kết phosphodiester giữa nhóm 5'-PO_4 của polynucleotide này với nhóm 3'-OH của polynucleotide khác.



Hình 9.5. Phương thức cơ bản để tạo dòng gen trong vi khuẩn *E. coli*

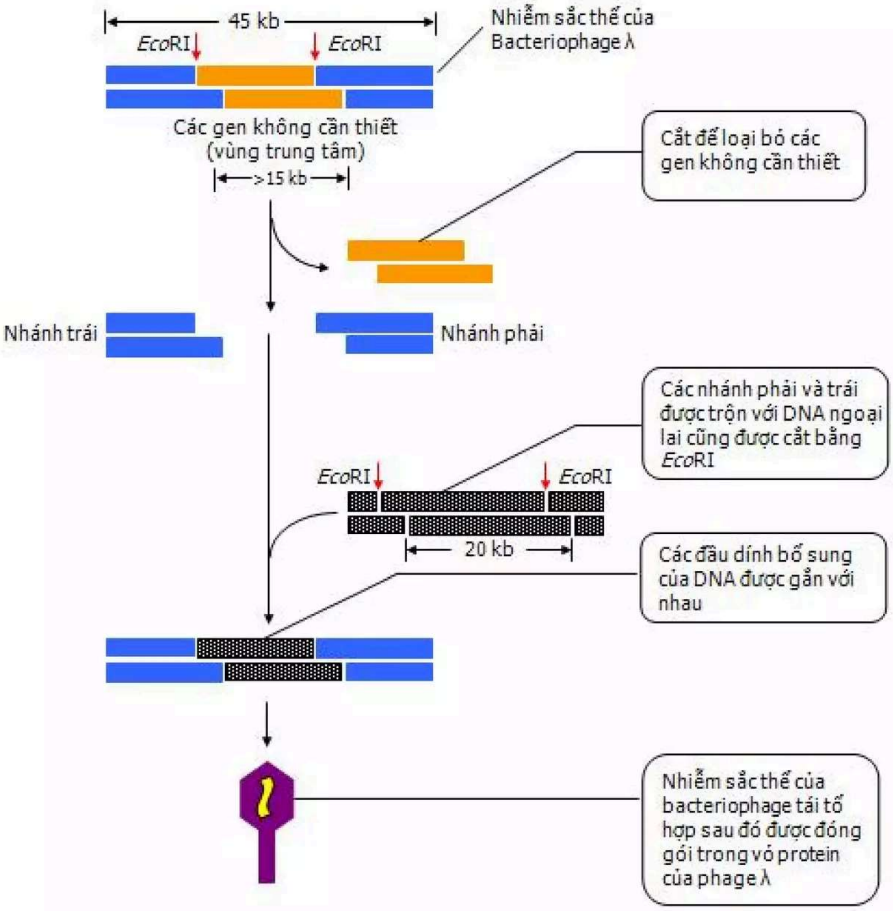
Hình 9.5 trình bày toàn bộ phương thức sản xuất DNA tái tổ hợp (tạo dòng gen). Plasmid được cắt ở các vị trí xác định bằng enzyme hạn chế. DNA của một genome ngoại lai được cắt bởi cùng một loại enzyme, một số đoạn trong đó có thể có gen quan tâm. Plasmid và các đoạn của genome được phối trộn và kết hợp nhờ enzyme DNA ligase. Các plasmid tái tổ hợp được biến nạp vào vi khuẩn bằng đồng nuôi cấy plasmid và vi khuẩn.

2. Bacteriophage } vector

Các bacteriophage có nhiều ưu điểm khi được sử dụng làm vector tạo dòng. Vector dược sử dụng rộng rãi nhất là bacteriophage λ để gây nhiễm vào tế bào *E. coli*. Một trong những ưu điểm chính của phage λ là có hiệu quả chuyển DNA vào trong tế bào vi khuẩn cao. Ưu điểm thứ hai là một phần ba của genome phage λ là không cần thiết cho sự xâm nhiễm và sinh sản của nó trong tế bào vật chủ, không có những gen này, một tiểu thể λ

vẫn chuyển thành công DNA của nó vào vi khuẩn và sinh sản được. Các gen không cần thiết này, khoảng >15 kb, có thể được thay thế bằng một đoạn DNA ngoại lai khoảng 15-23 kb. Ưu điểm thứ ba là DNA sẽ không bị đóng gói trong vỏ λ trừ khi nó dài từ 40-50 kb; vì thế các đoạn DNA ngoại lai không bao giờ được chuyển vào tế bào trừ khi chúng được chèn vào trong genome phage λ, điều kiện cần thiết để đảm bảo đoạn DNA ngoại lai sẽ được tái bản sau khi nó vào trong tế bào vật chủ.

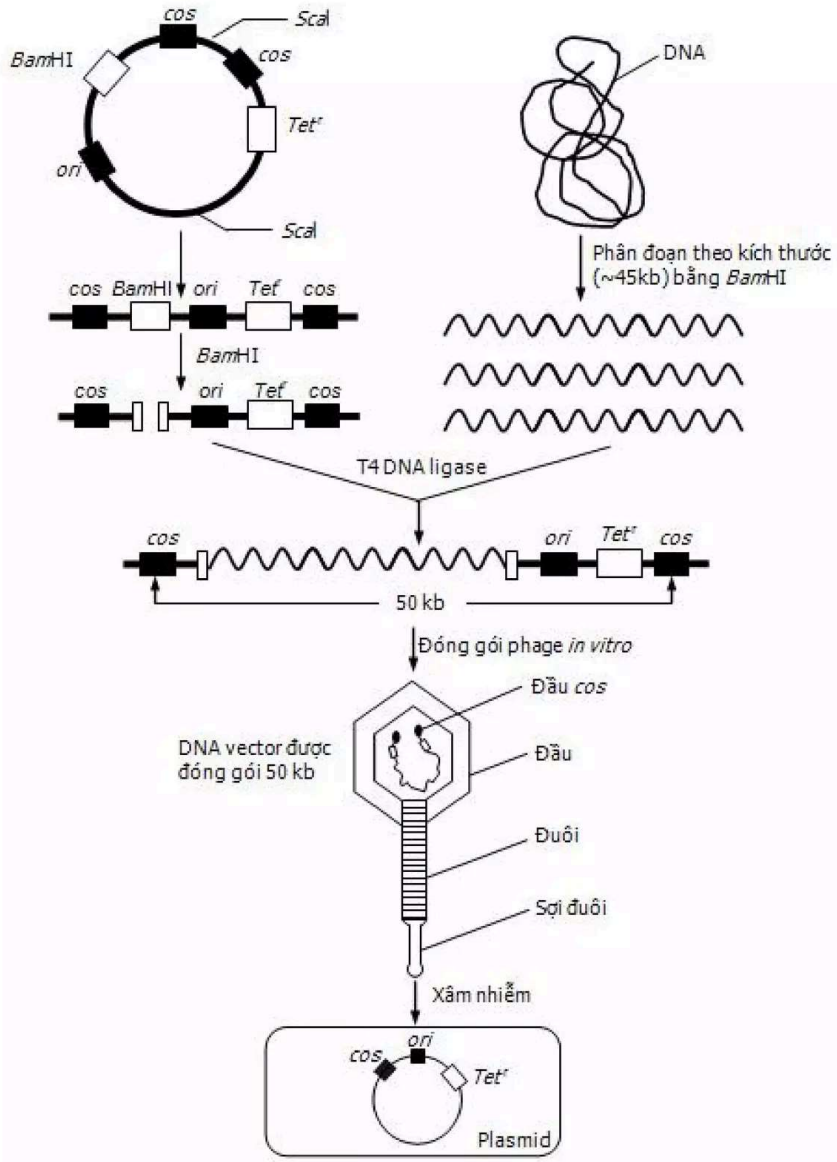
Các gen cần thiết của genome phage λ được định vị trong một cụm. Các chủng của phage λ, được gọi là vector thay thế, đã được biến đổi di truyền với một vị trí RE duy nhất cho *Eco*RI (chủng phage λ thế hệ mới EMBL 3 có ba vị trí cho các RE: *Eco*RI, *Bam*HI và *Sal*I) trên một mặt khác của các gen không cần thiết (Hình 9.6). Vì thế, có thể loại bỏ các gen không cần thiết bằng *Eco*RI. DNA ngoại lai được cắt bằng *Eco*RI sẽ có đầu dính bổ sung cho đầu của các đoạn DNA của phage λ cần thiết (nhánh trái và phải), để có thể được kết nối bằng DNA ligase. Genome của phage λ có các đoạn sợi đơn ngắn được gọi là các vị trí *cos* cần cho sự đóng gói DNA trong đầu của phage. Genome tái tổ hợp của phage λ sau đó có thể được đóng gói trong vỏ protein và được bổ sung vào trong *E. coli*. Các phage λ gây ra sự xâm nhiễm DNA tái tổ hợp của chúng vào trong tế bào vật chủ, nơi nó sẽ được tái bản. Chỉ các đoạn DNA có kích thước thích hợp và mang các gen cần thiết được đóng gói trong các vỏ của phage, cung cấp một hệ thống chọn lọc tự động cho các vector tái tổ hợp.



Hình 9.6. Bacteriophage λ là một vector tạo dòng hiệu quả

3. Cosmid vector

Các vector phage λ chỉ có thể mang các đoạn DNA có kích thước khoảng 15-23 kb. Tuy nhiên, các cosmid vector lại mang được các đoạn DNA có kích thước lớn hơn nhiều, khoảng 45 kb.



Hình 9.7. Tạo dòng trong cosmid. Hai vị trí *cos* gần vị trí cắt hạn chế *Sca*I và *Bam*HI. DNA nguồn bị cắt bởi *Bam*HI để phân đoạn có kích thước khoảng 45 kb. Phân tử plasmid DNA được cắt bởi *Bam*HI và *Sca*I. Hai mẫu DNA này được trộn vào nhau và được gắn bằng T4 DNA ligase. Sau khi gắn xong, những phân tử này được đóng gói trong phần đầu của phage λ, và những phần tử có thể lây nhiễm sẽ được hình thành sau khi tạo đuôi.

Cosmid là các plasmid nhỏ mang các vị trí *cos* của phage; chúng có thể được đóng gói trong vỏ virus và được chuyển vào vi khuẩn nhờ sự xâm nhiễm của virus. Do tất cả các gen virus, ngoại trừ các vị trí *cos*, là không có, nên cosmid có thể mang các đoạn DNA ngoại lai lớn hơn hai lần các đoạn mà phage λ vector có thể mang. Các cosmid

vector có các thành phần sau: (1) một điểm khởi đầu sao chép của plasmid (*ori*); (2) một số các vị trí cắt hạn chế đơn; (3) một hoặc hơn các gen chỉ thị chọn lọc; và (4) các vị trí *cos* cho phép đóng gói DNA trong đầu của phage.

DNA ngoại lai được chèn vào trong cosmid trong cùng một phương thức với plasmid: cosmid và DNA ngoại lai đều được cắt bởi cùng một RE để tạo ra các đầu bổ sung (đầu dính), và chúng được liên kết với nhau bằng DNA ligase. Các cosmid tái tổ hợp được hợp nhất trong vỏ, và các tiểu thể phage được sử dụng để xâm nhiễm vào tế bào vi khuẩn, ở đó cosmid sẽ tái bản như một plasmid. Bảng 9.2 so sánh các tính chất của các vector plasmid, phage λ và cosmid.

Bảng 9.2. So sánh các vector plasmid, phage } và cosmid

Vector tạo dòng	Kích thước DNA có thể được tạo dòng	Phương pháp sinh sản	Phương pháp chuyển vào vi khuẩn
Plasmid	Khoảng 10 kb	Tái bản plasmid	Biến nạp
Phage λ	Khoảng 20 kb	Sinh sản phage	Xâm nhiễm phage
Cosmid	Khoảng 45 kb	Sinh sản plasmid	Xâm nhiễm phage

4. Thư viện cDNA

Thư viện cDNA (complementary DNA) là tập hợp các đoạn DNA bổ sung (cDNA) được tổng hợp từ mRNA của một bộ phận trong cơ thể sinh vật. Sử dụng thư viện cDNA có hai ưu điểm sau:

- Các dòng cDNA chứa trình tự mã hóa liên tục của một gen. Nhiều gen ở eukaryote là gián đoạn, có chứa nhiều đoạn intron. Sau khi cắt mRNA tiền thân (pre-mRNA) và nối lại, các đoạn intron đã bị loại và mRNA hoàn chỉnh (mature mRNA) có trình tự mã hóa liên tục được tạo thành. Do cDNA được tạo thành từ khuôn mẫu mRNA hoàn chỉnh nên các dòng cDNA có thể tổng hợp protein cần thiết với số lượng lớn như mong muốn.

- Nhiều protein được tổng hợp với số lượng lớn do những tế bào chuyên hóa và như vậy trong các tế bào này mRNA của protein đó sẽ có tỷ lệ cao và thư viện cDNA được tạo ra từ các tế bào này sẽ có nhiều cDNA mã hóa cho các protein tương ứng. Sự dồi dào cDNA của một vài loại nào đó đã làm giảm nhẹ đáng kể việc xác định đúng dòng mong muốn từ thư viện gen.

Phương pháp tổng hợp gen từ mRNA ngày càng được phát triển, nó kết hợp với các phương pháp khác của sinh học phân tử cho phép ứng dụng trong nhiều lĩnh vực.

Xây dựng một thư viện cDNA bao gồm năm bước chính sau:

- Tinh sạch mRNA từ RNA tổng số của một phận cơ thể sinh vật.

- Tổng hợp sợi cDNA thứ nhất từ khuôn mẫu mRNA nhờ enzyme phiên mã ngược (reverse transcriptase).

- Gây biến tính (đun sôi) thể lai mRNA-cDNA để phá hủy sợi mRNA bằng RNase H của *E. coli*. Tổng hợp sợi cDNA thứ hai từ khuôn mẫu sợi cDNA thứ nhất nhờ enzyme DNA polymerase với primer là vòng cặp tóc của nó để thu được phân tử cDNA sợi đôi.

- Cắt vòng cặp tóc bằng enzyme nuclease S1 và dùng enzyme Klenow sửa chữa hai đầu của sợi đôi cDNA để tạo ra đầu bằng.

- Gắn các đoạn nối (linker) vào hai đầu của cDNA sợi đôi trước khi tạo dòng trong vector thích hợp để xây dựng thư viện cDNA.

5. Thư viện genomic DNA

Thư viện genomic DNA là một tập hợp các đoạn DNA cùng đại diện cho một genome nguyên vẹn (hoặc gần nguyên vẹn) của cá thể mà DNA được bắt nguồn từ đó. Các thư viện được dùng chủ yếu hoặc để sàng lọc theo các phương pháp khác nhau nhằm phân lập một (hoặc các) trình tự nucleotide quan tâm đặc biệt, hoặc để xác định vị trí và thứ tự của các trình tự trong genome mà từ đó thư viện được xây dựng (physical mapping).

Xây dựng một thư viện genomic DNA bao gồm bốn bước chính: Cắt DNA, gắn DNA vào vector, đóng gói các dòng tái tổ hợp trong vỏ protein của virus để làm bước trung gian trước khi xâm nhập vào tế bào vật chủ, và chuyển cấu trúc đó vào tế bào vật chủ.

Thư viện genomic DNA có nhiều ứng dụng, chẳng hạn để lập bản đồ vật lý (physical mapping) của DNA và xác định các gen gây bệnh hoặc các chuỗi DNA quan tâm cho những phân tích khác.

Sản xuất ra các dòng mang các đoạn chèn DNA khác nhau nhưng gối lên nhau có nhiều thuận lợi và thư viện có thể được sử dụng trong quá trình chromosome walking. Chromosome walking thường được thực hiện với thư viện của cosmid, phage λ hoặc YAC^2.

Các thư viện genomic DNA cũng cần thiết cho việc xác định các gen gây bệnh bằng cách tạo dòng chức năng (functional cloning). Theo hướng này, thông tin về chức năng của gen được khai thác để phân lập gen mong muốn từ thư viện. Một oligonucleotide có trình tự dựa trên chuỗi amino acid từng phần được dùng như là một probe (mẫu dò) để phân lập dòng cDNA bằng cách sàng lọc thư viện cDNA. Dòng cDNA này sau đó có thể được dùng để sàng lọc thư viện genome nhằm phân lập các dòng genomic DNA và cho phép khảo sát đặc điểm của chuỗi genomic hoàn chỉnh.

IV. Biểu hiện gen ngoại lai trong vi khuẩn

Về mặt lý thuyết, kỹ thuật DNA tái tổ hợp cho phép đưa bất kỳ một gen nào đó từ một sinh vật này vào một sinh vật khác. Vấn đề quan trọng là làm sao để gen ngoại lai có thể biểu hiện trong cơ thể vật chủ.

Để biểu hiện tất cả các gen ngoại lai trong *E. coli* phải bắt đầu bằng việc gắn đoạn gen ngoại lai vào trong vector biểu hiện (thường là plasmid). Các vector biểu hiện là vector có thể mang các gen ngoại lai mong muốn cho phép thực hiện sự phiên mã của các bản sao được tạo dòng và sự dịch mã các mRNA của chúng trong *E. coli*. Vector này phải có đủ các cấu trúc cần thiết sau:

- Các trình tự mã hóa gen chỉ thị (marker) để đảm bảo duy trì vector trong tế bào.
- Một promoter kiểm soát phiên mã (ví dụ: *lac, trp* hoặc *tac*) cho phép sản xuất một lượng lớn mRNA từ các gen được tạo dòng.
- Các trình tự kiểm soát dịch mã như vùng liên kết ribosome được bố trí thích hợp và codon khởi đầu AUG.
- Một polylinker để đưa gen ngoại lai vào trong một hướng chính xác với promoter.

Chỉ khi được cấu trúc đầy đủ như thế, các vector biểu hiện mang gen ngoại lai mới được biến nạp vào chủng *E. coli* thích hợp. Nếu đoạn gen ngoại lai không nằm giữa promoter và vị trí kết thúc phiên mã, thì nó sẽ không được phiên mã.

1. Các protein nguyên thể tái tổ hợp

Các protein nguyên thể (native protein) có thể được sản xuất trong *E. coli* bằng cách sử dụng promoter mạnh và một vùng liên kết ribosome (ribosome binding sites-RBS) hiệu quả. Để biểu hiện gen prokaryote có RBS mạnh chỉ cần cung cấp một promoter là đủ. Trong khi đó, để biểu hiện một gen eukaryote (hoặc một gen prokaryote với một RBS yếu) cần phải cung cấp cả promoter lẫn RBS.

1.1. Biểu hiện của gen prokaryote-Promoter

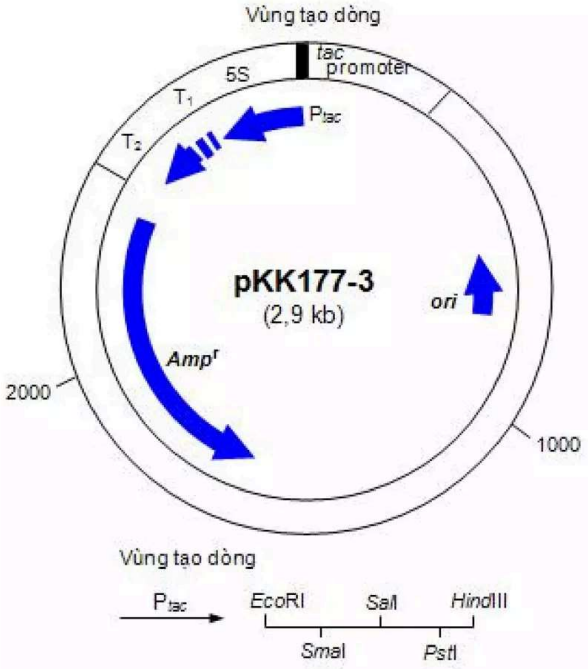
Bước đầu tiên khi biểu hiện các protein của eukaryote trong vi khuẩn là chọn một vector biểu hiện mang promoter mạnh của prokaryote. Các promoter thích hợp nhất cho biểu hiện của gen ngoại lai ở *E. coli* là những promoter có khả năng điều chỉnh mạnh. Điển hình là promoter (lai) *trp-lac*.

Promoter *trp-lac* còn gọi là promoter *tac* (một dạng promoter lai giữa promoter *trp* và promoter *lac*) đã được sử dụng thành công để sản xuất một lượng lớn protein trong *E. coli* (Hình 9.8). Promoter *trp* được điều chỉnh bởi gen ức chế *trp* và có thể được cảm ứng bởi sự bổ sung của 3-indolylacetic acid (3-IAA) vào môi trường hoặc bằng cách thiếu tryptophan. Trong khi đó, promoter *lac* được điều chỉnh bởi gen ức chế *lac* và vì thế có thể được cảm ứng nhờ bổ sung nhân tố cảm ứng isopropyl-β-D-thiogalactoside (IPTG) vào môi trường nuôi cấy vi khuẩn. Cuối cùng, promoter (lai) *trp-lac* mang đoạn *trp*-35

được gắn với đoạn *lac*-10 và operator *lac* được điều chỉnh bởi gen ức chế *lac* (*lac* repressor).

1.2. Biểu hiện của gen eukaryote-Promoter và vùng liên kết ribosome

Để có mức độ biểu hiện cao của gen ngoại lai trong *E. coli*, không những phải sử dụng các promoter mạnh để sản xuất một lượng lớn mRNA mà còn sử dụng các RBS để đảm bảo chắc chắn mRNA được dịch mã một cách có hiệu quả. Trong *E. coli*, RBS bao gồm một mã khởi đầu AUG và một trình tự nucleotide ngược hướng từ codon khởi đầu. Trình tự này được gọi là chuỗi Shine-Dalgarno (SD) giàu A-G, bổ sung cho đầu 3' của 16S rRNA của *E. coli*. Liên kết của ribosome với mRNA được khởi đầu bằng cách bắt cặp giữa chuỗi SD trong mRNA và trình tự ở đầu 3' của 16S rRNA.



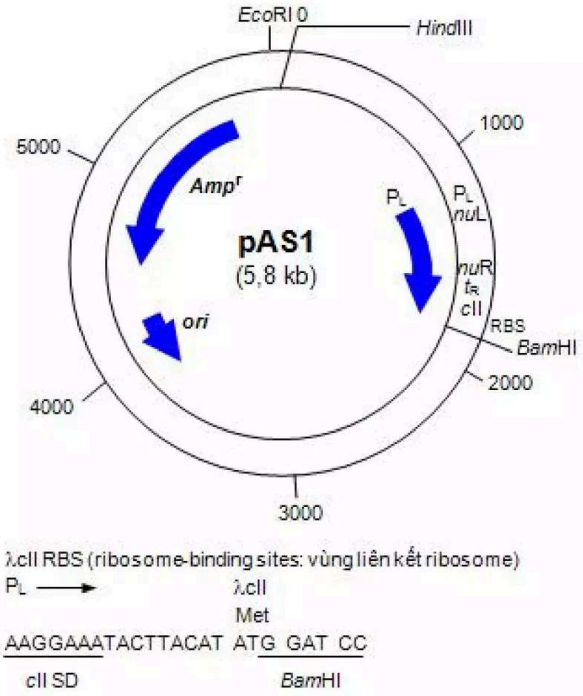
Hình 9.8. Vector pKK177-3. pKK177-3 là một *tac* vector chứa vùng tạo dòng gen ngoại lai cùng hướng với promoter *tac*. Cùng hướng với vùng này là *rrn*B mang gen 5S của *E. coli* và hai nhân tố kết thúc phiên mã T_1 và T_2.

Hiệu suất dịch mã của mRNA chịu sự chi phối của một số yếu tố như sau :

- Mức độ bổ sung giữa chuỗi SD và đầu 3' của 16S rRNA.
- Không gian và khả năng để chuỗi DNA nằm giữa chuỗi SD và codon AUG.

- Nucleotide theo sau AUG ảnh hưởng sự liên kết ribosome.

Đưa codon ATG vào trong gen được tạo dòng và duy trì RBS của vi khuẩn bằng cách dùng vector pAS1 (Hình 9.9). Vector pAS1 mang promoter P_L *v*à RBS của gen *c*II của bacteriophage λ, codon khởi đầu ATG được dung hợp trực tiếp với phần mã hóa đầu amino của gen eukaryote muốn biểu hiện. Cách làm này có thể hạn chế việc tối ưu khoảng cách giữa chuỗi SD của vi khuẩn và codon khởi đầu ATG của gen eukaryote để có được sự biểu hiện hiệu quả của gen. Đoạn DNA có gắn promoter và chuỗi SD sau đó được biến nạp vào các chủng *E. coli* thích hợp. Sàng lọc thể biến nạp để thu các khuẩn lạc có mức độ phiên mã cao.



Hình 9.9. Vector pAS1. Vector pAS1 là một plasmid dài khoảng 5,8 kb mang promoter P_L của bacteriophage λ và vị trí cắt hạn chế duy nhất *Bam*HI định vị ở codon khởi đầu ATG của gen *c*II của bacteriophage λ.

2. Các protein dung hợp tái tổ hợp

2.1. Protein dung hợp

Protein dung hợp còn gọi là protein lai được mã hóa bởi một gen lai (fusion gene) do sự dung hợp *in vitro* các đoạn gen khác nhau. Vì vậy, protein dung hợp sẽ mang trình tự amino acid của hai protein khác biệt. Các protein dung hợp được xây dựng cho các

mục đích khác nhau. Sự dung hợp gen được thực hiện bằng cách gắn phần mã hóa của gen được tạo dòng ở gần đầu tận cùng 3' của gen *lacZ*. Protein dung hợp có những ưu điểm chính sau:

- Protein dung hợp thường được sản xuất với hàm lượng lớn do sự khởi đầu phiên mã và dịch mã được điều khiển bởi các trình tự tiêu chuẩn của *E. coli*.

- Dung hợp các trình tự ngoại lai với các gen của *E. coli* thường cho kết quả các sản phẩm ổn định hơn các protein ngoại lai nguyên thể.

2.2. Vector biểu hiện các gen dung hợp với gen *lacZ*

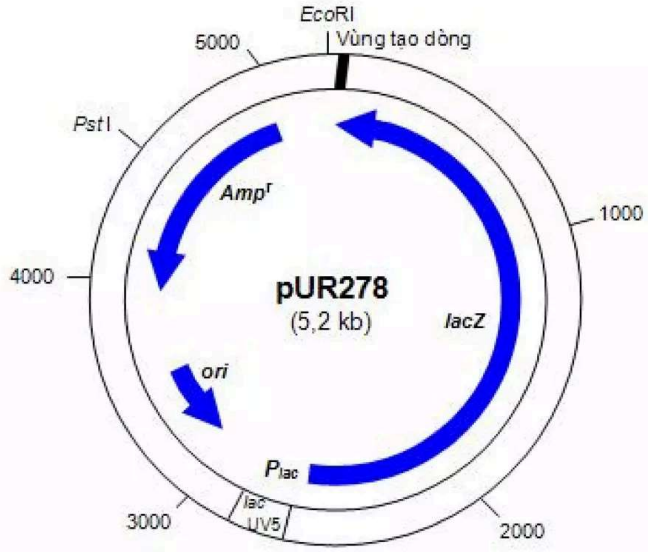
Một số hệ thống vector được phát triển để biểu hiện các gen dung hợp với gen *lac*Z, điển hình là các vector họ pUR (Hình 9.10). Chọn vector và vị trí cắt hạn chế thích hợp người ta có thể tiến hành dung hợp cho hầu hết gen được tạo dòng.

2.3. Phát hiện các protein dung hợp

Gắn vector plasmid (ví dụ: pUR) thích hợp với đoạn DNA ngoại lai để tạo ra một sự dung hợp trong khung đọc. Biến nạp các vector này vào *E. coli*. Kiểm tra các khuẩn lạc riêng biệt mang đoạn DNA ngoại lai mong muốn bằng cách tách chiết DNA của vector plasmid, sau đó cắt bằng enzyme hạn chế thích hợp và điện di kiểm tra trên agarose gel. Sàng lọc các khuẩn lạc sản xuất protein dung hợp.

2.4. Tách chiết các protein dung hợp để sản xuất kháng thể

Protein dung hợp có thể được tách chiết theo một số cách: dùng dịch chiết urea, sắc ký ái lực aminophenylthiogalactoside, điện di sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) hoặc phối hợp giữa các cách này. Kỹ thuật đơn giản nhất là SDS-PAGE, nhuộm gel và phát hiện băng protein mới, sau đó cắt băng này ra khỏi gel, đông khô rồi nghiền thành bột mịn. Bột này sẽ được dùng để tiêm vào thỏ để sản xuất kháng thể.



Vùng tạo dòng

lacZ

pUR278 TGT CAA AAA GGG GAT CCG TCG ACT CTA GAA AGC TTA TCG ATG
BamHI *SalI* *XbaI* *HindIII* *ClaI*

pUR288 TGT CGG GGA TCC GTC GAC TCT AGA AAG CTT ATC GAT GAT
BamHI *SalI* *XbaI* *HindIII* *ClaI*

pUR289 TGT CAG GGG ATC CGT CGA CTC TAG AAA GCT TAT CGA TGA
BamHI *SalI* *XbaI* *HindIII* *ClaI*

pUR290 TGT CAA AAA GGG GAT CCG TCG ACC TGC AGC CAA GCT TAT CGA TGA
BamHI *SalI* *PstI* *HindIII* *ClaI*

pUR291 TGT CGG GGA TCC GTC GAC CTG CAG CCA AGC TTA TCG ATG
BamHI *SalI* *PstI* *HindIII* *Cla I*

pUR292 TGT CAG GGG ATC CGT CGA CCT GCA GCC AAG CTT ATC GAT
BamHI *SalI* *PstI* *HindIII* *ClaI*

Hình 9.10. Các vector họ pUR. Đây là các vector dung hợp với gen *lacZ,* có các vị trí tạo dòng *Bam*HI, *Sal*I, *Pst*I, *Xba*I, *Hin*dIII và *Cla*I ở đầu 3' của gen *lacZ*. Sự chèn vào của đoạn DNA ngoại lai (cDNA) trong các vị trí tạo dòng thích hợp cho phép sản xuất protein dung hợp của β-galactosidase hoạt động với chuỗi peptide được mã hóa bởi DNA ngoại lai.

3. Xác định mức độ biểu hiện của gen được tạo dòng

Nói chung, có ba cách thường được dùng để xác định mức độ biểu hiện protein ngoại lai của gen được tạo dòng:

- Điện di polyacrylamide gel có SDS để xác định protein có kích thước thích hợp được sản xuất ở mức độ cao trong các tế bào mang vector biểu hiện. Thông thường, protein quan tâm có thể quan sát bằng cách nhuộm gel với Coomassie Brilliant Blue hoặc bằng $AgNO_3$. Nếu không thấy có băng protein mới khi dùng các thuốc nhuộm này, thì đánh dấu sự trao đổi chất với 100 μCi của [^{35}S]Met hoặc [^{35}S]Cys trên 1 mL dịch nuôi

cấy trong 5 phút. Sau đó, sử dụng kỹ thuật SDS-PAGE hoặc phóng xạ tự ghi có thể phát hiện được protein quan tâm.

- Phân tích Western blot bằng cách dùng các kháng thể liên kết đặc hiệu với protein quan tâm đã được thẩm tích lên màng nitrocellulose sau khi thực hiện điện di SDS-polyacrylamide gel.

- Nếu mức độ biểu hiện thấp thì nên đặt gen *lac*Z cùng hướng với gen được biểu hiện. Như vậy, nếu sự phiên mã hoặc dịch mã hạn chế biểu hiện thì những thay đổi trong hệ thống biểu hiện có thể được kiểm soát bằng những thay đổi trong hoạt tính của β-galactosidase.

▪ Phân tích Western blot

- Kỹ thuật SDS-PAGE

Là kỹ thuật điện di trên polyacrylamide gel với sự có mặt của SDS cho phép phân tách các phân tử protein có khối lượng khác nhau. SDS có điện tích âm rất lớn và có khả năng liên kết với mạch peptide. Như vậy, số lượng SDS tương tác với protein tỷ lệ với kích thước phân tử protein và điện tích của SDS bám vào có thể làm bất cứ phân tử protein nào cũng chuyển động trong điện trường từ cực âm sang cực dương. Do đó, bằng phương pháp điện di, có thể phân tách riêng biệt các phân tử protein có khối lượng phân tử khác nhau.

- Phản ứng lai kháng nguyên-kháng thể

Phản ứng kháng nguyên-kháng thể có tính đặc hiệu rất cao. Vì vậy, có thể áp dụng phản ứng này để phát hiện sự có mặt và tinh sạch protein. Kháng thể (antibody) được sản xuất khi đưa kháng nguyên vào động vật thí nghiệm (thỏ, chuột…) và được tinh sạch từ máu động vật sau khi gây nhiễm. Những kháng thể tạo ra bằng cách này là những kháng thể đa dòng (polyclonal antibodies-do các tế bào lympho khác nhau tiết ra), do đó chúng có khả năng nhận biết một số kháng nguyên. Ngược lại, kháng thể đơn dòng (monoclonal antibodies) chỉ tương tác với một kháng nguyên nhất định.

Kháng thể được đánh dấu bằng enzyme (hoặc bằng chất phát huỳnh quang) để phát hiện protein đặc hiệu (thường được thẩm tích/chuyển thấm lên màng nitrocellulose sau khi chạy điện di SDS, và cố định ở đó) thông qua kỹ thuật Western blot (hoặc immunoblot) (Hình 9.11). Sau khi protein trên màng lai gắn với kháng thể thứ nhất đặc hiệu và tiếp đến là kháng thể thứ hai có đánh dấu enzyme (alkaline phosphatase, horse-radish peroxidase…) thì phức hợp này sẽ được liên kết với cơ chất để tạo màu. Sự hiện diện của protein ngoại lai (sản phẩm dịch mã của gen ngoại lai được chuyển vào tế bào vật chủ) sẽ được phát hiện nhờ sự xuất hiện màu của phản ứng lai.

Sự phân bố của protein đặc hiệu trong tế bào và tổ chức mô cũng có thể phát hiện bằng kỹ thuật lai *in situ* (*in situ* hybridization) với nguyên tắc tương tự Western blot. Ngoài ra, kháng thể được sử dụng để tinh sạch protein đặc hiệu bằng kết tủa miễn dịch

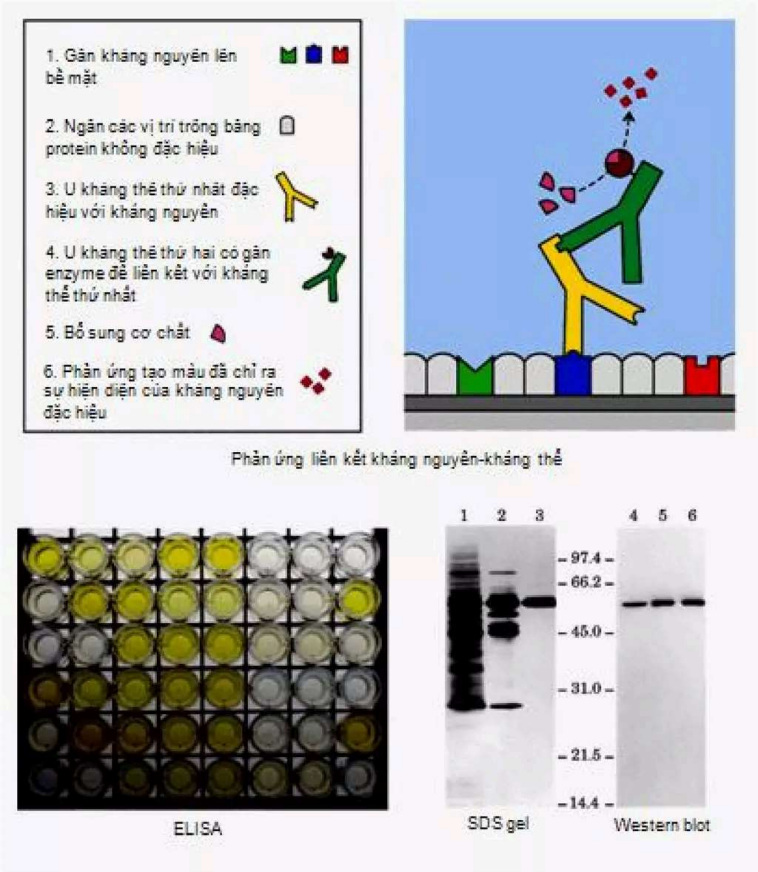
hoặc sắc ký ái lực (affinity chromatography). Kháng thể đánh dấu được dùng để định lượng kháng nguyên trong kỹ thuật xét nghiệm hấp thụ miễn dịch liên kết enzyme (enzyme-linked immunosorbent assay) gọi tắt là ELISA (Hình 9.11).

V. Phương pháp phát hiện dòng vi khuẩn có DNA tái tổ hợp

1. Lai khuẩn lạc và vết tan

DNA được tạo dòng trong plasmid hoặc YAC sản xuất ra các khuẩn lạc khi các nuôi cấy biến nạp được dàn mỏng trên đĩa agar chứa môi trường sinh trưởng và nuôi cấy dưới những điều kiện thích hợp. Các bacteriophage sinh tan tế bào bị chúng xâm nhiễm và sản xuất ra các plaque (vết tan) có dạng hình tròn chu vi khoảng 2-3 mm có màu sáng trên thảm vi khuẩn (bacterial lawn) của lớp agar đỉnh (trong trường hợp này người ta chuẩn bị đĩa agar có hai lớp: lớp agar đỉnh có nồng độ agar thấp để bacteriophage dễ sinh tan tế bào vi khuẩn, và lớp agar đáy có nồng độ agar cao hơn).

Các vector, như mô tả ở trên, thường chứa gen chỉ thị cho phép chọn lọc những tế bào vật chủ mang vector (thể biến nạp). Các chỉ thị này thường là gen kháng kháng sinh và các tế bào biến nạp sinh trưởng trên môi trường chứa kháng sinh tương ứng.



Hình 9.11. Kỹ thuật Western blot và ELISA dựa trên phản ứng liên kết kháng nguyên-kháng thể

Ngoài ra, các vector còn chứa các gen chỉ thị bổ sung để phân biệt các tế bào biến nạp chứa đoạn chèn của DNA ngoại lai với các tế bào chứa các vector tự tái tạo lại vòng. Ví dụ: gen *lacZ* mã hóa enzyme β-galactosidase.

Một dòng (clone) chứa các chuỗi DNA quan tâm đặc biệt có thể được xác định bởi lai khuẩn lạc hoặc vết tan. Một lượng nhỏ khuẩn lạc biến nạp hoặc vết tan được chuyển lên màng nitrocellulose hoặc nylon bằng cách phủ (overlay) màng này lên trên đĩa agar. DNA được biến tính và cố định trên màng bằng đun nóng (baking) hoặc chiếu tia cực tím (UV-crosslinking), và sau đó được lai trong đệm chứa probe đã đánh dấu đồng vị phóng xạ có trình tự bổ sung một phần của chuỗi được xác định. Ví dụ: Có thể một oligonucleotide tổng hợp nhân tạo, có nguồn gốc từ genomic DNA từng phần, cDNA hoặc trình tự protein hoặc sản phẩm PCR. Trong một vài trường hợp, probe có thể được thiết kế dựa trên trình tự bắt nguồn từ gen tương đồng của các loài khác và thường được lai với cường lực thấp. Các probe thừa được rửa khỏi màng để ủ với phim X-quang. Theo hướng của phim (sau khi rửa) so sánh với đĩa agar gốc, sau đó sẽ có khả năng gắn kết các khuẩn lạc thực tế với các khuẩn lạc lai dương tính tương ứng trên phim X-quang.

2. Sàng lọc thư viện gen bằng PCR

PCR (polymerase chain reaction) cũng có thể được ứng dụng để sàng lọc các thư viện genomic DNA hoặc cDNA được xây dựng trong các plasmid vector hoặc bacteriophage vector. Ưu điểm chính của sàng lọc bằng PCR so với sàng lọc dựa trên lai truyền thống là ít tốn thời gian hơn. Sàng lọc bằng PCR có thể được đảm bảo trong 3-4 giờ trong khi đó nó có thể mất một vài ngày trước khi phát hiện bằng kỹ thuật lai phân tử (molecular hybridization). Kỹ thuật sàng lọc bằng PCR thể hiện chỉ thị về kích thước của các đoạn chèn DNA được tạo dòng hơn là trình tự của đoạn chèn, tuy nhiên các primer PCR đặc trưng cho đoạn chèn DNA ngoại lai cũng có thể được sử dụng. Điều này cho phép khảo sát chính xác hơn đặc điểm của các dòng từ thư viện cDNA và genomic DNA.

▪ Nguyên tắc của PCR

PCR là một kỹ thuật được sử dụng phổ biến trong công nghệ sinh học hiện đại và đã đóng góp rất lớn cho những tiến bộ về sinh học phân tử, đánh dấu một bước tiến vô cùng quan trọng tương đương với việc khám phá ra các enzyme hạn chế và kỹ thuật Southern blot.

Kỹ thuật PCR dựa trên cơ sở phản ứng mở rộng primer nhờ enzyme Taq polymerase để khuếch đại *in vitro* các nucleic acid đặc trưng. PCR cho phép khuếch đại theo hàm mũ lên đến hàng triệu lần các đoạn DNA có chiều dài khoảng từ 200-3.000 bp. Đoạn DNA được khuếch đại (DNA đích) được nhận diện nhờ cặp primer đặc trưng (oligonucleotide) thường có chiều dài khoảng 20 nucleotide.

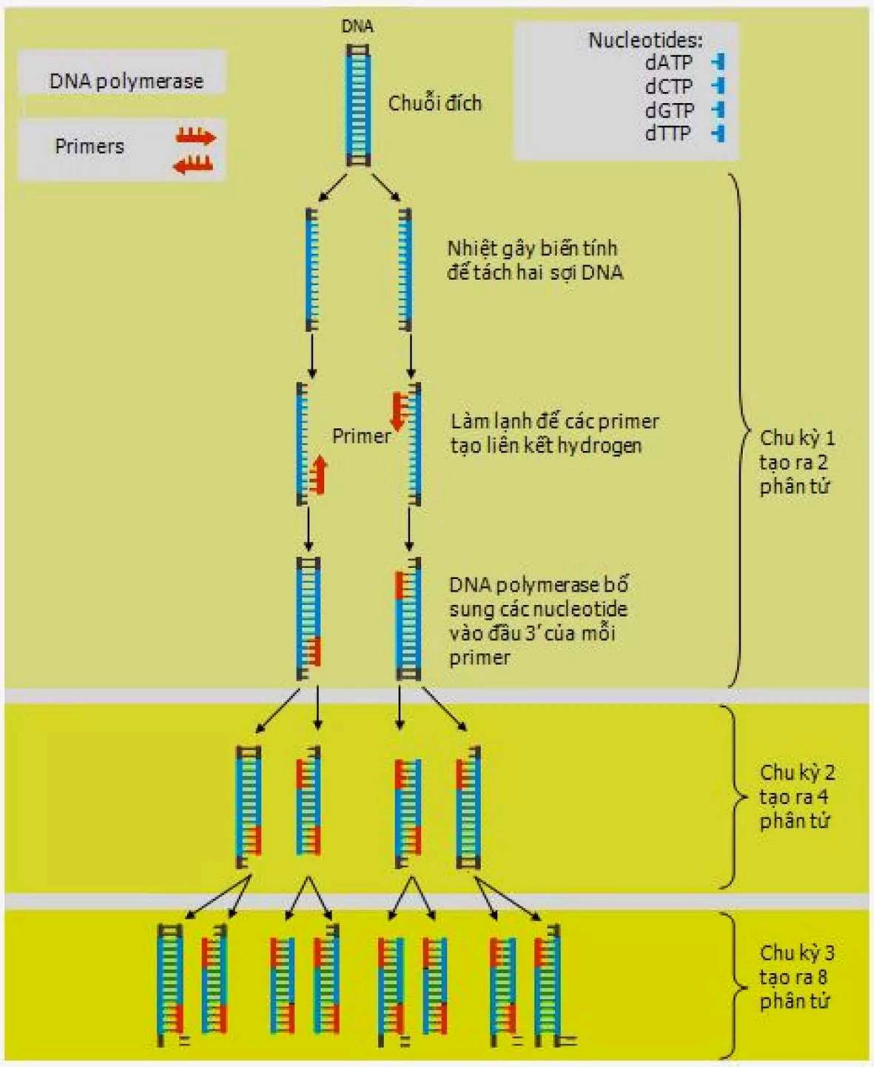
Taq polymerase là một loại enzyme DNA polymerase chịu nhiệt (có ở vi khuẩn chịu nhiệt độ cao *Thermus aquaticus*) được dùng để tổng hợp các đoạn DNA mới trong môi trường có 4 loại deoxyribonucleotide (dATP, dCTP, dGTP và dTTP) và hai primer, trên cơ sở khuôn mẫu của một đoạn DNA nhất định đã biết hoặc chưa biết trình tự. Các đoạn DNA mới hình thành lại được sử dụng làm khuôn mẫu. Sau nhiều chu kỳ, số lượng đoạn DNA nói trên được nhân lên gấp nhiều lần, nhờ vậy có thể đủ số lượng để tách ra, phân tích trình tự hoặc tạo dòng... Primer ở bên trái tác động trên sợi DNA 3'-5' được gọi là primer thuận (forward primer-ký hiệu là F). Primer ở bên phải tác động trên sợi DNA 5'-3' được gọi là primer ngược (reverse primer-ký hiệu là R).

Nguyên tắc của PCR được trình bày trong hình 9.12. Theo đó, từ chu kỳ thứ hai Taq polymerase bắt đầu tạo ra các đoạn DNA có chiều dài xác định. Các primer thường là một oligonucleotide tổng hợp (synthetic oligonucleotide) có khoảng 10-20 nucleotide hoặc dài hơn. Nếu biết trình tự của đoạn gen cần khuếch đại thì có thể tổng hợp nhân tạo các primer tương ứng để thực hiện PCR và tách chúng ra bằng kỹ thuật điện di. PCR thường tiến hành khoảng 25-35 chu kỳ, qua đó từ 10^{-6} µg DNA ban đầu có thể khuếch đại (amplification) lên tới trên 1 µg (khoảng 2 kb). Mỗi chu kỳ PCR bao gồm ba giai đoạn có nhiệt độ khác nhau:

- *Gây biến tính (denaturation) ở 90-95^oC*. Trong giai đoạn biến tính phân tử DNA khuôn mẫu ở dạng xoắn kép được tách thành hai sợi đơn (single strands).

- *Gắn primer (annealing) ở 40-65^oC*. Trong giai đoạn này các primer gắn vào các vị trí có trình tự tương đồng ở DNA khuôn mẫu. Các liên kết ion tạo thành một đoạn nhỏ (các primer đã lắp ráp chính xác) và trên các đoạn nhỏ DNA sợi đôi đó (khuôn mẫu và primer) enzyme Taq polymerase có thể bắt đầu quá trình sao chép khuôn mẫu.

- *Kéo dài phân tử (extension) ở 70-72^oC*. Đây là khoảng nhiệt độ tối thích cho Taq polymerase tiến hành tổng hợp DNA bắt đầu từ các vị trí có primer theo chiều 5'→3'.



Hình 9.12. Sơ đồ phản ứng chuỗi DNA polymerase (PCR)

3. Sàng lọc các thư viện cDNA bằng các probe khác nhau

Sàng lọc thư viện cDNA bằng cách lai nucleic acid là phương pháp được sử dụng phổ biến và đáng tin cậy nhất để tìm kiếm các dòng quan tâm. Sàng lọc bằng phương pháp lai nucleic acid cho phép phân tích đồng thời nhiều dòng và nhanh, không đòi hỏi các dòng cDNA phải hoàn chỉnh và sản phẩm được tổng hợp trong tế bào vật chủ phải có hoạt tính sinh học hoặc kháng nguyên. Hơn nữa, cơ sở lý thuyết của kỹ thuật lai nucleic acid đã được hiểu biết đầy đủ. Điều này cho phép phát triển một số lượng lớn các kỹ thuật khác nhau có thể điều chỉnh các probe của nucleic acid có các chiều dài và đặc điểm khác nhau.

- **Các probe tương đồng.** Các probe tương đồng chứa ít nhất một phần của chuỗi nucleic acid chính xác của dòng cDNA quan tâm. Chúng được dùng trong nhiều trường hợp khác nhau, ví dụ: khi một dòng bộ phận của cDNA hiện có được sử dụng để phân lập một dòng hoàn chỉnh từ thư viện cDNA. Thông thường, đoạn bắt nguồn từ một đầu hoặc đầu kia của dòng hiện có được phân lập, đánh dấu phóng xạ *in vitro* và dùng để thăm dò thư viện. Lai với các probe tương đồng thường được tiến hành dưới các điều kiện nghiêm ngặt.

- Các probe tương đồng từng phần. Các probe tương đồng từng phần được dùng để phát hiện các dòng cDNA có quan hệ họ hàng, nhưng không giống hệt nhau hoàn toàn, với các trình tự của probe.

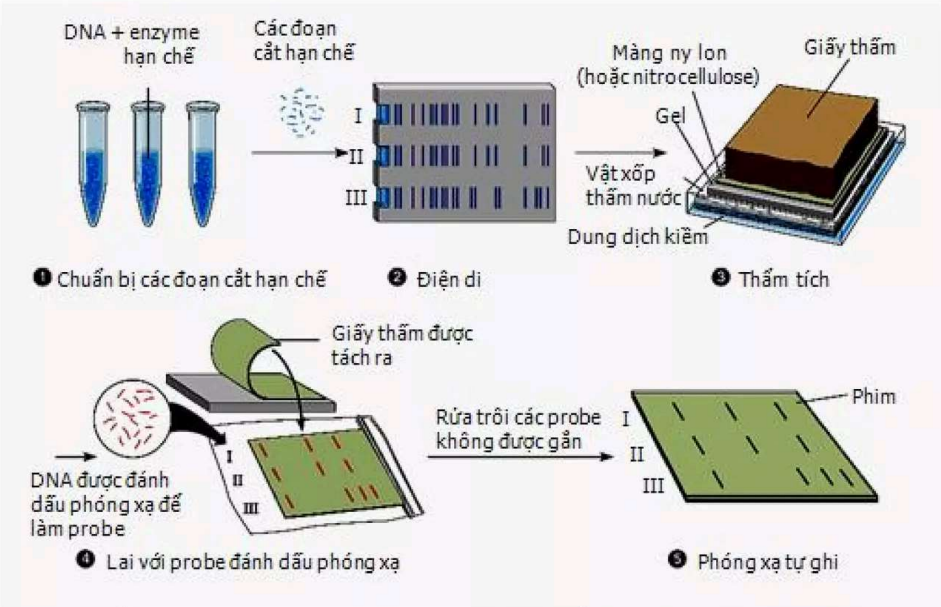
- Các probe oligonucleotide nhân tạo. Các probe oligonucleotide nhân tạo là các vùng dNTP của trình tự xác định được tổng hợp *in vitro*. Trình tự của các probe này được suy luận, bằng cách dùng mã di truyền, từ các vùng ngắn của chuỗi amino acid đã biết của protein quan tâm. Do sự thoái biến của mã di truyền[3], chuỗi amino acid đề xuất có thể không được đặc trưng chính xác bởi các oligonucleotide đơn được dự đoán cho trình tự. Mặc dù, trong rất nhiều trường hợp, trình tự các amino acid giống nhau có thể được đặc trưng bởi nhiều oligonucleotide khác nhau.

4. Phân tích genomic DNA bằng phương pháp lai Southern

Việc phát hiện và xác định các chuỗi nucleic acid đặc trưng là công việc thường xuyên trong nghiên cứu sinh học phân tử. Nguyên tắc của kỹ thuật này là dựa vào sự lai phân tử, dưới các điều kiện thích hợp hai chuỗi nucleic acid đơn tạo thành một phân tử lai. Phản ứng lai phụ thuộc rất nhiều vào mức độ tương đồng của hai chuỗi nucleotide. Việc tạo thành phân tử sợi đôi như thế xảy ra chủ yếu thông qua liên kết hydrogen giữa các base guanosine với cytosine, và adenosine với thymidine. Thành phần và sự phân bố của các base không giống rõ rệt với các phân tử nucleic acid cho kết quả các tính chất lai khác nhau, vì thế liên kết hydrogen (hoặc lai phân tử giữa các base tương đồng) khi được ghép cặp thích hợp đã cung cấp một công cụ có giá trị để xác định các chuỗi liên quan hoặc đồng nhất với chuỗi nucleic acid quan tâm (probe).

Trong số các kỹ thuật khác nhau sử dụng lai phân tử để phân tích nucleic acid, thì kỹ thuật được sử dụng phổ biến nhất là lai mẫu dò nucleic acid được đánh dấu với nucleic acid đích được cố định trên một vật đỡ rắn, đặc trưng là màng nitrocellulose hoặc nylon.

Trình tự của kỹ thuật lai Southern blot bao gồm các bước sau: phân tách các đoạn cắt hạn chế của DNA hệ gen bằng điện di agarose gel, chuyển DNA từ agarose gel lên màng lai, lai các mẫu dò được đánh dấu đồng vị phóng xạ với các nucleic acid được cố định trên màng lai (Hình 9.13).



Hình 9.13. Sơ đồ của kỹ thuật lai Southern blot. Các mẫu DNA đích và genomic DNA (A) và (B) trong ví dụ này được cắt bằng *Eco*RI và được phân đoạn bằng điện di trên agarose gel. DNA của plasmid mang đoạn chèn DNA dùng làm mẫu dò (C) cũng được cắt bằng *Eco*RI và điện di như là một đối chứng dương tính.

Các thí nghiệm lai trên màng thường được tiến hành bằng cách đánh dấu probe bằng ^{32}P (mặc dù các đồng vị phóng xạ khác như ^{35}S cũng có thể được sử dụng). Phản ứng lai của Southern đòi hỏi hoạt tính đặc hiệu của probe tối thiểu phải là 10^9 dpm/µg, mặc dù hoạt tính 10^8 dpm/µg có thể được xem như là chấp nhận được trong các ứng dụng cần cường lực thấp. Các phương pháp không dùng đồng vị phóng xạ (ví dụ: digoxigenin-dUTP) ngày càng được sử dụng phổ biến hơn mặc dù độ nhạy của chúng là không lớn. Nhưng các kỹ thuật không dùng đồng vị phóng xạ an toàn hơn cho nghiên cứu viên, tạo ra các probe có thể được bảo quản trong thời gian dài trước khi dùng và kết quả phát hiện các tín hiệu lai nhanh hơn.

VI. Các ứng dụng của công nghệ DNA tái tổ hợp

Công nghệ DNA tái tổ hợp hiện nay có rất nhiều ứng dụng trong đời sống. Các ứng dụng này bao gồm sản xuất dược phẩm và các loại hóa chất khác, các vi khuẩn đặc biệt, các cây trồng nông nghiệp quan trọng, và các vật nuôi đã được biến đổi di truyền... Kỹ thuật này cũng được sử dụng rộng rãi trong các xét nghiệm y khoa và (trong một vài trường hợp) đang được dùng để kiểm tra các khiếm khuyết di truyền ở người. Hàng trăm công ty hiện nay đang đặc biệt phát triển các sản phẩm thông qua biến đổi di truyền các cơ thể sống. Dưới đây là các ứng dụng chính của công nghệ DNA tái tổ hợp.

1. Ứng dụng trong dược phẩm

Sản phẩm thương mại đầu tiên được phát triển bằng công nghệ DNA tái tổ hợp là các dược phẩm dùng trong điều trị các bệnh và các rối loạn ở người. Năm 1979, Tổ hợp Eli Lilly bắt đầu sản xuất thương mại insulin người bằng công nghệ DNA tái tổ hợp. Trước thời điểm này, tất cả insulin dùng trong điều trị bệnh tiểu đường được tách chiết từ tụy của các vật nuôi cho thịt. Mặc dù nguồn insulin này có tác dụng tốt cho nhiều người mắc bệnh tiểu đường, nhưng nó vẫn không phải là insulin người, và một số người đã trải qua các phản ứng phụ với protein ngoại lai. Gen insulin người sau đó đã được chèn vào plasmid và chuyển vào vi khuẩn để sản xuất insulin người. Các dược phẩm được sản xuất thông qua công nghệ DNA tái tổ hợp bao gồm: hormone sinh trưởng người (cho trẻ em mắc bệnh còi), nhân tố tạo tơ huyết (chữa bệnh khó đông máu), hoạt tố plasminogen mô (dùng để hòa tan các huyết khối trong bệnh nhân bị nhồi máu cơ tim), interferon, interleukin, DNA vaccine…

2. Các vi khuẩn đặc biệt

Các vi khuẩn có vai trò quan trọng trong các quá trình công nghiệp, bao gồm sản xuất ethanol từ các nguyên liệu thực vật, lọc khoáng từ quặng, xử lý nước thải và các loại chất thải khác. Các vi khuẩn đang sử dụng trong các quá trình này được biến đổi di truyền bằng công nghệ DNA tái tổ hợp sao cho chúng có thể hoạt động hiệu quả hơn. Các chủng vi khuẩn mới hữu ích thu được từ kỹ thuật hiện đại này đang được phát triển để phân hủy các hóa chất độc và các chất gây ô nhiễm, tăng cường thu hồi dầu, tăng nitrogen cho cây trồng, và ức chế sinh trưởng của các vi khuẩn và vi nấm gây bệnh.

3. Các sản phẩm nông nghiệp

Công nghệ DNA tái tổ hợp cũng đã có một ảnh hưởng quan trọng đến sản xuất nông nghiệp, nơi mà hiện nay nó được dùng để tạo ra các cây trồng và vật nuôi mang các đặc điểm có giá trị. Trong nhiều năm, các nhà bệnh học thực vật đã thừa nhận rằng thực vật bị nhiễm bệnh virus thể nhẹ sau đó sẽ kháng lại sự nhiễm trùng các chủng mang độc tính mạnh. Vì vậy, các nhà di truyền học đã tạo ra tính kháng virus trong cây trồng bằng

cách chuyển các gen vỏ protein của virus vào tế bào thực vật. Cây bí (squash) biến đổi di truyền có tên là Freedom II, mang các gen của virus 2 gây bệnh khảm ở dưa hấu (watermelon mosaic virus 2) và virus gây bệnh khảm màu vàng ở cây zucchini (zucchini yellow mosaic virus) để chống sự nhiễm trùng virus.

Một hướng khác được biến đổi di truyền đó là tính kháng sâu (pest) ở thực vật đã làm giảm sự phụ thuộc vào các thuốc trừ sâu hóa học. Độc tố protein từ vi khuẩn *Bacillus thuringiensis* đã giết một cách chọn lọc các ấu trùng của các sâu bọ côn trùng nhất định nhưng lại không gây độc cho sự sống của các động vật hoang dã, người và các loài côn trùng khác. Gen độc tố được phân lập từ vi khuẩn, liên kết với promoter hoạt động sau đó được chuyển vào trong ngô, cà chua, khoai tây và bông. Gen này đã sản xuất độc tố đối với côn trùng trong cây, và sâu bướm khi ăn cây đã bị chết.

Công nghệ DNA tái tổ hợp cũng đã cho phép phát triển tính kháng chất diệt cỏ ở thực vật. Một vấn đề quan trọng trong nông nghiệp là điều khiển các cỏ dại cạnh tranh nước, ánh sáng và chất dinh dưỡng với cây trồng. Mặc dù chất diệt cỏ có tác dụng giết cỏ dại, nhưng chúng cũng có thể gây nguy hiểm cho cây trồng. Các gen cung cấp tính kháng cho nhiều loại chất cỏ dại đã được chuyển vào cà chua, đậu tương, bông, cây cải dầu và các cây trồng thương mại quan trọng khác. Khi đồng ruộng có canh tác các cây trồng này được phun thuốc diệt cỏ, thì cỏ dại bị giết nhưng các cây trồng biến đổi di truyền lại không bị ảnh hưởng. Năm 1999, đã có hơn 21 triệu ha cây đậu tương biến đổi di truyền và 11 triệu ha cây ngô chuyển gen được trồng trên thế giới.

Công nghệ DNA tái tổ hợp cũng được ứng dụng cho các vật nuôi. Ví dụ: gen sản xuất hormone sinh trưởng được phân lập từ gia súc và được tạo dòng trong *E. coli,* các vi khuẩn này sản xuất một lượng lớn hormone sinh trưởng của bò, yếu tố liên quan đến tăng sản xuất sữa. Các động vật chuyển gen được phát triển để mang gen mã hóa cho các sản phẩm dược liệu. Ví dụ: một gen người mã hóa cho nhân tố VIII (hình thành tơ huyết) được liên kết với vùng điều hòa của gen sản xuất β-lactoglobulin ở cừu, một protein tạo sữa. Gen dung hợp được tiêm vào phôi cừu, tạo ra một con cừu chuyển gen có khả năng sản xuất trong sữa của nó nhân tố đông máu của người. Một phương thức tương tự được dùng để chuyển gen α_1-antitrysin, một protein dùng để điều trị bệnh nhân khí thũng di truyền (hereditary emphysema), vào trong cừu. Cừu cái mang gen này sản xuất rất nhiều α_1-antitrysin khoảng 15g trong 1 lít sữa của chúng.

4. Thuốc oligonucleotide

Một ứng dụng công nghệ DNA tái tổ hợp trong thời gian gần đây đó là phát triển các thuốc oligonucleotide có trình tự phân tử DNA hoặc RNA nhân tạo ngắn có thể được dùng để điều trị bệnh. Các antisense oligonucleotide bổ sung cho các RNA không mong muốn, như là RNA của virus. Khi được bổ sung vào tế bào, thì các antisense DNA này liên kết với mRNA của virus và ức chế sự dịch mã của chúng.

Các DNA oligonucleotide sợi đơn liên kết chặt chẽ với các trình tự DNA khác, tạo thành một phân tử triplex DNA. Sự tạo thành triplex DNA cản trở liên kết của RNA polymerase và các protein khác cần cho sự phiên mã. Các oligonucleotide khác là các ribozyme, các phân tử RNA mà chức năng như là các enzyme. Những phức hợp này liên kết với các phân tử mRNA đặc hiệu và cắt chúng thành từng đoạn, phá hủy khả năng mã hóa protein của chúng. Một số thuốc oligonucleotide đã sẵn sàng để thử nghiệm trong điều trị bệnh AIDS và ung thư.

5. Liệu pháp gen

Đối với loại bệnh di truyền, bệnh do đột biến gen người ta phải cần đến sự can thiệp của liệu pháp gen, một hướng chữa bệnh gắn liền với các kỹ thuật tiên tiến trong lĩnh vực công nghệ sinh học hiện đại như các vi thao tác gen, sửa đổi thay thế gen...

Liệu pháp gen (gene therapy) thực chất là phương pháp chữa bệnh bằng gen. Có nhiều khái niệm khác nhau về liệu pháp gen, nhưng cách hiểu chung nhất là tập hợp các biện pháp để sử dụng các gen cần thiết (còn gọi là gen trị liệu) nhằm mục đích chữa bệnh cho con người.

Trong đó, gen trị liệu có thể là:

- Các gen hoạt động bình thường (gen lành) có thể đưa vào tế bào để thay thế gen hỏng, gen mất chức năng, khôi phục hoạt động bình thường của tế bào và sự sống của cơ thể.

- Là những gen có khả năng mã hóa một protein đặc hiệu, khi đưa vào tế bào sống có thể tạo nên các loại protein đặc hiệu. Các loại protein đặc hiệu có thể ức chế hoạt động của một gen khác trong tế bào, kìm hãm khả năng phân chia của tế bào hoặc gây chết các tế bào bị bệnh.

- Những gen khi đưa vào tế bào hoạt động đồng thời với các gen bệnh (gen bị đột biến trong tế bào) làm hạn chế tác động của gen bệnh hoặc hỗ trợ, bù đắp cho các gen bị hỏng.

- Gen trị liệu còn là các gen bất hoạt được đưa vào tế bào thay thế cho một gen lành nào đó, nhằm hạn chế sản phẩm không cần thiết của gen lành hoặc tạo ra cho tế bào một trạng thái mới, có tác dụng chống lại bệnh tật.

- Gen trị liệu có thể là những đoạn oligonucleotide có tác dụng kìm hãm hoạt động của gen bị hỏng, bị đột biến trong tế bào.

6. Chẩn đoán bệnh để can thiệp sớm

Bằng kỹ thuật chọc ối, kết hợp đồng thời phân tích máu bố mẹ bằng enzyme hạn chế, người ta có thể chẩn đoán sớm trước khi sinh (vì enzyme hạn chế có khả năng phân biệt được gen đột biến với gen bình thường). Các đoạn DNA cắt ra, được phân tách qua phương pháp điện di, đem lai với các mẫu dò DNA hoặc RNA đã đánh dấu bằng đồng vị phóng xạ ^{32}P hoặc huỳnh quang. Ảnh phóng xạ tự ghi hoặc tín hiệu huỳnh quang cho ta

thấy các đoạn DNA được lai với các mẫu dò. Các đoạn này được tách ra để nghiên cứu và xác định đột biến hay bình thường bằng enzyme hạn chế, vì một số đột biến di truyền có ảnh hưởng đến các vị trí dành cho enzyme hạn chế.

Ứng dụng các kỹ thuật sinh học phân tử trong lĩnh vực này có nhiều thuận lợi hơn các phương pháp kinh điển. Thứ nhất, có thể rút ngắn được thời gian nhờ ứng dụng kỹ thuật PCR, chẳng hạn chẩn đoán bệnh HIV-1, dùng phương pháp PCR chỉ mất một ngày so với 3-4 tuần nuôi cấy của phương pháp truyền thống. Thứ hai, đối với các tác nhân gây bệnh khó nuôi cấy hay không thể nuôi cấy thì kỹ thuật sinh học phân tử là giải pháp duy nhất, chẳng hạn *Chlamydia*, *Brucella,* viêm gan B,... Thứ ba, việc tạo dòng một gen dùng làm probe đơn giản hơn việc sản xuất kháng thể đặc hiệu, hơn nữa với một probe người ta có thể phát hiện tất cả các kiểu huyết thanh (serotype) của tác nhân gây bệnh, điều mà chỉ một kháng thể không thể làm được. Cuối cùng, nếu trước kia phải cần đến nhiều kỹ thuật (quan sát dưới kính hiển vi, nuôi cấy trên một loạt môi trường, miễn dịch học...) thì hướng chẩn đoán bằng sinh học phân tử chỉ cần một kỹ thuật (lai phân tử hoặc PCR).

Đối với tác nhân là virus, các kỹ thuật lai phân tử và PCR đã cho phép chẩn đoán nhiều nhóm như papillomavirus, HBV, HIV. Đặc biệt phương pháp lai tại chỗ còn cho phép xác định trực tiếp virus trên lát cắt sinh thiết.

Đối với các tác nhân vi khuẩn, đã có nhiều bộ kit chẩn đoán sinh học phân tử được thương mại hóa như: *Mycobacterium tuberculosis, Mycobacterium pneumoniae, Salmonella, Streptococcus pneumoniae, Haemophilus influenzae*...

Đối với các tác nhân ký sinh trùng, người ta cũng đã thành công trong chẩn đoán *Plasmodium, Schistosoma, Trypanosoma, Toxoplasma, Leishmania*… bằng lai phân tử. Phương pháp PCR còn cho phép phát hiện tác nhân ngay trong vật trung gian truyền bệnh.

7. DNA fingerprinting

Sử dụng các trình tự DNA để nhận dạng một người, được gọi là DNA fingerprinting, là một công cụ rất mạnh để khảo sát tội phạm và các ứng dụng pháp lý khác.

Trong một ứng dụng đặc trưng, DNA fingerprinting có thể được dùng để xác nhận một sự nghi ngờ hiện diện ở hiện trường gây án. Một mẫu DNA của máu, tóc, tinh dịch hoặc mô của cơ thể được thu thập từ hiện trường. Nếu mẫu là rất nhỏ, phương pháp PCR có thể được dùng để khuếch đại cho đủ DNA, sẵn sàng cho việc kiểm tra. Các mẫu DNA bổ sung được thu thập từ một hoặc nhiều nghi can khác.

Mỗi một mẫu được cắt với một hoặc nhiều RE, và kết quả các đoạn DNA được phân tách bằng điện di. Các đoạn trên gel được biến tính và chuyển lên màng nitrocellulose bằng thẩm tích Southern (blot). Một hoặc nhiều probe (được đánh dấu phóng xạ) sau đó được lai với màng nitrocellulose và được phát hiện bằng phóng xạ tự ghi. Kiểu của băng được tạo ra bởi DNA từ mẫu được thu thập ở hiện trường sau đó được so sánh với các kiểu được tạo ra bởi DNA từ những nghi can.

Các probe được dùng trong DNA fingerprinting phát hiện các vùng thay đổi cao của genome, vì thế cơ hội DNA từ hai người sản xuất chính xác cùng một kiểu băng là rất thấp. Khi một số probe được dùng trong phân tích, thì xác suất mà hai người có cùng một tập hợp các kiểu là gần như không có (trừ khi họ là cặp sinh đôi giống hệt nhau).

Mẫu khớp nhau từ hiện trường và nghi can có thể cung cấp bằng chúng rằng nghi can có mặt tại hiện trường vụ án.

Probe được dùng phổ biến nhất trong DNA fingerprinting bổ sung cho các trình tự lặp lại ngắn được tìm thấy rất nhiều trong genome người. Con người khác nhau rất nhiều về số lượng bản sao của các trình tự lặp lại này; vì vậy, các đa hình này được gọi là số lượng biến thiên của các đoạn lặp lại tuần tự (VNTRs).

DNA fingerprinting cũng được sử dụng để cung cấp thông tin về mối quan hệ huyết thống và các nguồn cơ thể sống khác. Ví dụ: DNA fingerprinting được dùng để xác định một số mẫu vi khuẩn gây bệnh than (anthrax) gửi bằng bưu điện đến những người khác trong năm 2001 tất cả là từ một nguồn mẫu.

8. Lập bản đồ gen

Một đóng góp có ý nghĩa của công nghệ DNA tái tổ hợp đó là cung cấp một lượng lớn các marker (chỉ thị) di truyền được dùng trong lập bản đồ gen. Một nhóm trong số các marker được dùng trong lập bản đồ gen đó là restriction fragment length polymorphism (RFLP). RFLP là những biến đổi đa hình trong các kiểu của các đoạn DNA được tạo ra khi phân tử DNA được cắt cùng một enzyme hạn chế. Nếu DNA từ hai người được cắt bởi cùng một enzyme và kiểu của các đoạn DNA khác nhau được tạo ra, thì hai người này phải có sự khác nhau về các trình tự DNA của họ. Những sự khác nhau này được di truyền và có thể được dùng trong lập bản đồ, tương tự với phương thức mà trong đó sự khác nhau về allele được dùng để lập bản đồ các gen thông thường.

Theo truyền thống, bản đồ gen được dựa vào việc sử dụng các sai khác di truyền tạo ra những sai khác kiểu hình có thể quan sát dễ dàng. Đáng tiếc là do hầu hết các tính trạng bị ảnh hưởng bởi đa gen và môi trường, nên số lượng các tính trạng có cơ sở di truyền đơn giản thích hợp cho việc sử dụng bản đồ bị giới hạn. RFLP cung cấp một số lượng lớn các marker di truyền có thể được dùng để lập bản đồ gen của sinh vật.

Tài liệu tham khảo/đọc thêm

1. Dale JW and Von Schanzt M. 2002. From Gene to Genome. *John Wiley & Sons, Ltd.* West Sussex, UK.

2. Erlich HA. 1989. PCR Technology: Principles and Applications for DNA Amplification. *Stockton Press*, New York, USA.

3. Glick BR and Pasternak JJ. 2003. Molecular Biotechnology: Principles and Applications of Recombinant DNA. 3rd ed. *ASM Press*, USA.

4. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

5. Maniatis T, Fritsch EF and Sambrook J. 1989. Molecular Cloning-A Laboratory Manual. *Cold Spring Habor Laboratory*, USA.

6. Primrose SB, Twyman R and Old RW. 2001. Principles of Gene Manipulation. 6th ed. *Blackwell Science*, Oxford, UK.

7. Rapley R and Walker JM. 1998. Molecular Biomethods Handbook. *Humana Press Inc*. New Jersey, USA.

8. Surzycki S. 2000. Basic Techniques in Molecular Biology. *Springer-Verlag*, Berlin, Heidelberg, Germany.

9. Walker JM and Rapley R. 2000. Molecular Biology and Biotechnology. *Chapman & Hall Limited*, London, UK.

[1]Chất tẩy làm biến đổi bề mặt tế bào để giải phóng các thành phần tế bào ra môi trường bên ngoài.
[2]YAC (yeast artificial chromosome): Nhiễm sắc thể nhân tạo của nấm men.
[3]Sự thoái biến (suy biến) của mã di truyền (degeneracy): Bình thường có thể có 64 tổ hợp bộ ba mã hóa khác nhau của 4 nucleotide trong khi đó chỉ có 20 amino acid phổ biến và một amino acid lại có thể có nhiều bộ ba mã hóa.

Phụ lục

Một số thuật ngữ cơ bản

Adenine (A). Một trong bốn nitrogen base có trong thành phần của DNA và RNA, bắt cặp với thymine (T) trên DNA và với uracil (U) trên RNA. Adenine cũng có trong thành phần của một số coenzyme.

Adenosine diphosphate (ADP). Một ribonucleoside 5'-diphosphate được cấu tạo từ adenine, đường ribose (5C) và hai gốc phosphate. ADP có tác dụng nhận phosphate trong chu trình năng lượng của tế bào.

Adenosine triphosphate (ATP). Một ribonucleoside 5'-triphosphate được cấu tạo từ adenine, đường ribose (5C) và ba gốc phosphate. ATP là phân tử chứa năng lượng hóa học chính của tế bào, chủ yếu được tập hợp trong ty thể (mitochondria) và lạp thể (chloroplast). Các gốc phosphate của ATP có mang các liên kết khi bị thủy phân sẽ phóng thích một năng lượng tự do lớn. Năng lượng của quá trình hô hấp hoặc quang hợp được sử dụng để tạo thành ATP từ ADP. Sau đó, ATP được biến đổi ngược trở lại thành

ADP ở nhiều vùng khác nhau của tế bào, năng lượng phóng thích ra được dùng để điều khiển các phản ứng hóa sinh nội bào. Đôi khi cũng xảy ra sự thủy phân tiếp ADP thành những AMP (adenosine monophosphate) để phóng thích năng lượng nhiều hơn.

Allele. Một trong nhiều dạng khác nhau của một gen chiếm một locus xác định trên nhiễm sắc thể. Các allele khác nhau của một gen có trình tự các nucleotide khác nhau, cùng liên quan đến một tính trạng của tế bào hoặc cơ thể. Ví dụ: trong thí nghiệm lai đậu Hà Lan của Mendel, màu vàng và màu xanh của hạt đậu là hai allele của gen quy định màu sắc hạt đậu. Cá thể được gọi là đồng hợp tử về một gen khi hai allele đó giống nhau, và gọi là dị hợp tử khi hai allele đó khác nhau. Ở trạng thái dị hợp tử allele này (allele trội) thường lấn át hiệu quả của allele kia (allele lặn). Allele quy định trạng thái bình thường là trội, trong khi các allele đột biến thường là lặn.

Amber codon. Là một trong ba codon (mã bộ ba UAG) kết thúc một trình tự của mRNA trong sinh tổng hợp protein.

Amino acid. Là một phân tử nhỏ mang một gốc amine ($-NH_3$) và một gốc carboxyl (-COOH) liên kết với cùng một nguyên tử carbon. Amino acid là đơn vị cấu trúc cơ sở của chuỗi polypeptide. Có 20 amino acid khác nhau trên các chuỗi polypeptide có trong tự nhiên. Trình tự sắp xếp của các amino acid trên chuỗi polypeptide quyết định cấu trúc và chức năng của polypeptide và protein mà nó tạo thành.

Aminoacyl-tRNA (aminoacyl-tRNA). Một ester giữa aminoacyl với tRNA tương ứng của amino acid đó.

AMP vòng (cyclic AMP, cAMP). Là một phân tử AMP mà trong đó gốc phosphate được liên kết với vị trí 3' và cả 5' của đường ribose. cAMP là thông tin thứ hai ở trong tế bào. Sự tạo thành cAMP (xúc tác bởi enzyme adenylate cyclase) được kích thích bởi một số hormone.

Ampicillin (Amp). Chất kháng sinh bán tổng hợp được dùng trong môi trường chọn lọc để chọn các tế bào mang đột biến khuyết dưỡng hoặc chọn dòng tế bào (tái tổ hợp) mang đoạn DNA được tạo dòng.

Antisense. Trình tự không mã hóa của một phân tử DNA sợi đôi. Chuỗi cây (chuỗi ngược chiều) bổ sung với mRNA và làm khuôn mẫu cho sự tổng hợp mRNA.

Antisense molecule. Là một phân tử gắn đặc hiệu lên DNA hoặc mRNA làm dừng quá trình phiên mã hoặc dịch mã. Thường phân tử này có nguồn gốc từ DNA và RNA.

Át chế (epistasis). Tác dụng tương hỗ giữa các gen khi một gen cản trở sự biểu hiện kiểu hình của một gen không alelle với nó.

ATP-ase. Enzyme thủy phân ATP thành ADP và phosphate, thường phối hợp với sự biến đổi năng lượng.

ATP-synthetase. Enzyme tổng hợp ATP từ ADP và phosphate trong quá trình phosphoryl oxy hóa (oxidative phosphorylation) ở màng trong của ty thể.

Baculovirus. Loại virus đặc biệt gây nhiễm các tế bào côn trùng và tạo ra các thể vùi lớn trong các tế bào bị nhiễm.

Bản đồ cắt hạn chế (restriction map). Trình tự các vị trí nhận biết (recognition sites) của tất cả các enzyme hạn chế (restriction enzyme hay restriction endonuclease, RE) trên một phân tử DNA.

Bản đồ di truyền (genetic map). Sơ đồ về trình tự sắp xếp của các cấu tử di truyền và khoảng cách tương đối giữa chúng trên nhiễm sắc thể (ở eukaryote) hoặc trên sợi DNA (ở prokaryote). Nếu các cấu tử đó là các gen trên nhiễm sắc thể và khoảng cách giữa chúng được xác định bằng tần số trao đổi chéo thì đó là bản đồ nhiễm sắc thể hay bản đồ liên kết gen. Nếu các cấu tử là các điểm bị đột biến bên trong gen và khoảng cách giữa chúng cũng được xác định bằng tần số trao đổi chéo giữa các điểm nói trên thì đó là bản đồ gen. Nếu các cấu tử di truyền là các điểm có thể bị cắt bởi enzyme hạn chế trên phân tử DNA thì gọi là bản đồ cắt hạn chế (restriction map). Xem bản đồ cắt hạn chế.

Bazơ đồng đẳng (analog base). Chất hóa học có cấu trúc phân tử rất giống các base bình thường của DNA. Chúng có thể thay thế các nitrogen base bình thường trong DNA và hoạt động như một tác nhân đột biến. Trong lần sao chép tiếp theo của DNA, base đồng đẳng có thể bắt cặp sai với một base bình thường, tạo nên đột biến điểm. Ví dụ: base đồng đẳng của adenine (A) là 2-aminopurine (AP) có thể gắn vào DNA ở vị trí của adenine; trong lần sao chép tiếp đó có thể bắt cặp với cytosine (C), trong lần sao chép tiếp theo nữa C kết cặp với guanine (G). Như vậy đã diễn ra sự thay thế cặp A-T bằng cặp G-C.

Bazơ nitơ (nitrogen base). Loại phân tử cấu tạo nên nucleic acid (DNA và RNA). Các nitrogen base có trong nucleic acid là adenine, guanine, cytosine và thymine (trong phân tử DNA) hoặc uracil (trong phân tử RNA). Trình tự sắp xếp của chúng dọc theo phân tử nucleic acid đã tạo nên thông tin di truyền của cơ thể sinh vật.

Bắt cặp bổ sung (complementary base pairing). Sự kết hợp thành từng đôi giữa các nitrogen base nằm trên hai mạch đơn của chuỗi xoắn kép DNA-DNA, DNA-RNA hoặc RNA-RNA thông qua các mối liên kết hydrogen. Sự bắt cặp đó mang tính đặc hiệu: guanine bắt cặp với cytosine, còn adenine bắt cặp với thymine trên DNA hoặc uracil trên RNA.

Biến đổi hậu dịch mã (post-translational modification). Sự thay đổi các liên kết hóa trị xảy ra trong chuỗi polypeptide, sau khi chuỗi polypeptide tách khỏi ribosome và trước khi trở thành protein hoạt động thực sự.

Biến đổi mRNA (mRNA processing). Quá trình tạo thành mRNA hoàn chỉnh (mature mRNA) hoạt động chức năng từ sản phẩm phiên mã sơ cấp là mRNA tiền thân chưa hoàn chỉnh (pre-mRNA). Trong quá trình này các đoạn intron bị loại bỏ và các đoạn exon được nối lại với nhau.

Biến nạp (transformation). Là quá trình truyền DNA ngoại lai vào một tế bào nhận, chẳng hạn sphaeroplast hoặc protoplast, và có thể hợp nhất trong nhiễm sắc thể nhờ

sự tái tổ hợp tương đồng hoặc được biến đổi trong một đơn vị sao chép tự trị (autonomous replicon). Sự biến nạp có thể xuất hiện trong các điều kiện tự nhiên ở một số vi khuẩn (ví dụ: *Bacillus, Haemophilus, Neisseria* và *Streptococcus*), nhưng ở nhiều vi khuẩn (ví dụ: *E. coli*) và các cơ thể sinh vật eukaryote sự biến nạp chỉ có thể xuất hiện ở những tế bào "thấm" được DNA bằng các phương pháp nhân tạo như: hóa biến nạp, điện biến nạp...

Biến tính (denaturation). Là hiện tượng chuyển từ dạng mạch kép sang dạng mạch đơn của DNA và RNA thường do nhiệt gây nên. Biến tính của protein là hiện tượng chuyển từ cấu hình hoạt động thành dạng không hoạt động.

Biểu hiện của gen (gene expression). Là các quá trình phiên mã (transcription) và dịch mã (translation) của một gen để tạo ra sản phẩm protein của nó.

Bộ khung (backbone). Thuật ngữ được dùng để chỉ chuỗi xoắn của khung đường-phosphate trong phân tử DNA mạch kép.

Cảm ứng (induction). Liên quan đến khả năng tổng hợp một số enzyme (sản phẩm của gen) ở vi khuẩn chỉ khi có mặt cơ chất (substrate) của các enzyme này trong môi trường. Khi sử dụng cho khái niệm điều hòa biểu hiện của gen, từ này có nghĩa là sự khởi động quá trình phiên mã do tương tác giữa một chất cảm ứng (inducer) với protein điều hòa. Ví dụ: trường hợp operon lac ở *E. coli,* lactose đóng vai trò chất cảm ứng có tác dụng mở operon để các gen trong operon hoạt động (phiên mã và dịch mã).

CAP (catabolite gene activator protein). Protein hoạt hóa gen dị hóa. Một protein điều hòa đặc biệt, kiểm soát sự khởi đầu phiên mã các gen sản xuất các enzyme cần thiết cho một tế bào, để có thể sử dụng những "thức ăn" khác khi không có mặt của glucose.

Cặp bazơ (base pair, bp). Là liên kết A-T hoặc C-G trên một phân tử DNA mạch kép, và là đơn vị đo chiều dài của một phân tử DNA.

Cấu trúc bậc một (primary structure). Bao gồm trình tự các amino acid và những cầu nối disulfide trong chuỗi hoặc giữa các chuỗi polypeptide. Ở trường hợp nucleic acid, đó là trình tự của các nucleotide trên một chuỗi DNA hoặc RNA được nối với nhau nhờ liên kết phosphodiester.

Cấu trúc bậc hai (secondary structure). Là sự nối liền những amino acid từ đầu nọ (đầu tận cùng N) tới đầu kia (đầu tận cùng C) của một phân tử protein. Cũng như vậy, trên một trình tự nucleic acid là sự nối liền những nucleotide từ đầu nọ (đầu 5') tới đầu kia (đầu 3') của một chuỗi DNA.

Cấu trúc bậc ba (tertiary structure). Cấu trúc ba chiều dạng cuộn xoắn của phân tử protein hoặc chuỗi polynucleotide.

Cấu trúc bậc bốn (quaternary structure). Cấu trúc ba chiều của một protein có nhiều tiểu đơn vị (subunit), theo kiểu mà những tiểu đơn vị kết hợp khớp với nhau.

Cầu disulfide (disulfide bridge). Liên kết đồng hóa trị tạo thành giữa hai chuỗi polypeptide qua trung gian của một gốc cystine.

Chạc ba sao chép (replication fork). Cấu trúc hình chữ Y hình thành khi phân tử DNA mạch kép dãn xoắn để lộ ra hai sợi đơn làm khuôn mẫu cho tổng hợp DNA (sự tái bản).

Chất cảm ứng (inducer). Một hợp chất hóa học hoặc một tác nhân của môi trường có tác dụng xúc tiến quá trình phiên mã ở operon của vi khuẩn hoặc các phân tử có tác dụng sản xuất ra số lượng lớn các enzyme trong trao đổi chất.

Chất dị nhiễm sắc (heterochromatin). Bao gồm các vùng của hệ gen thường xuyên ở trạng thái cô đặc và không biểu hiện di truyền. Trạng thái này có thể vĩnh viễn hay tạm thời.

Chất nhiễm sắc (chromatin). Phức hợp DNA và protein trong nhân ở giai đoạn nghỉ của quá trình phân bào. Trong giai đoạn này không thể phân biệt từng nhiễm thể riêng lẻ. Tên gọi này bắt nguồn từ phản ứng với các màu nhuộm đặc trưng cho DNA của phức hợp.

Chất nhiễm sắc thật (euchromatin). Chất nhiễm sắc chứa những gen được phiên mã.

Chất ức chế (repressor). Sản phẩm protein của một gen ức chế nằm trong thành phần operon, có tác dụng bám vào vùng chỉ huy của operon làm đóng operon lại, do đó quá trình phiên mã không bắt đầu được và tất cả các gen cấu trúc (gen mã hóa) trong operon đều ngừng hoạt động.

Chromosome walking. Kỹ thuật này dùng để lập bản đồ nhiễm sắc thể từ tập hợp các đoạn DNA cắt hạn chế chồng lên nhau (overlapping). Bắt đầu từ một thư viện trong đó chứa các đoạn DNA nói trên đã được tạo dòng. Một đoạn DNA mang một gen đã biết được lựa chọn và sử dụng như một probe để nhận dạng (ví dụ: bằng cách lai khuẩn lạc) các đoạn khác, là các đoạn chồng lên nhau chứa cùng một gen. Sau đó, trình tự nucleotide của các đoạn này sẽ được phân tích và nhờ vậy có thể xác định được toàn bộ các đoạn của nhiễm sắc thể. Từ đó, bản đồ của một vùng đặc biệt sẽ được xây dựng dần dần.

Chu trình sinh tan (lylic cycle). Một kiểu chu trình sống của thực khuẩn thể (bacteriophage) khi nó xâm nhiễm vi khuẩn, điều khiển các hoạt động sinh sản và sinh trưởng bằng các gen của nó và sinh ra các bacteriophage thế hệ con, chui ra khỏi tế bào vi khuẩn sau khi phá vỡ tế bào đó.

Chu trình tiềm tan (lysogenic cycle). Là hiện tượng hệ gen của bacteriophage hiện diện ở trạng thái ổn định và không sinh tan trong tế bào vật chủ sống của nó. Các tế bào vật chủ có thể tiếp tục sinh trưởng và phân chia, và sự sao chép của hệ gen bacteriophage (prophage) được phối hợp với nhiễm sắc thể của vật chủ sao cho khi tế bào phân chia thì prophage cũng được chuyển vào trong cả hai tế bào con. Prophage được duy trì bằng cách hoặc hợp nhất trong nhiễm sắc thể vật chủ (ví dụ: bacteriophage λ, bacteriophage Φ105) hoặc như là một plasmid bên ngoài nhiễm sắc thể (ví dụ:

bacteriophage P1 và bacteriophage F116). Tế bào vật chủ có thể hoặc không thể biểu hiện ra một kiểu hình biến đổi.

Chuỗi contig (contiguous sequence). Một đoạn trình tự dài được hình thành từ một số các đoạn phân tử ngắn chồng lên nhau (overlapping).

Chuỗi khảm (concatemer). Phân tử DNA bao gồm nhiều đoạn cá biệt nối với nhau thông qua các đầu dính.

Chuỗi mã hóa (coding sequence). Đoạn phân tử DNA mang mã di truyền xác định để phiên mã thành mRNA và sau đó dịch mã thành chuỗi polypeptide.

Chuyển gen (transgenic). Quá trình chuyển một đoạn DNA ngoại lai (foreign DNA) bằng các kỹ thuật khác nhau (*Agrobacterium*, vi tiêm, bắn gen, xung điện...) vào một cơ thể vật chủ (vi sinh vật, động vật hoặc thực vật).

Cistron. Thuật ngữ cũ, là một đơn vị của DNA tương ứng với một gen mã hóa một chuỗi polypeptide hoặc một phân tử RNA.

Cordycepin. Là 3' deoxyadenosine, một chất ngăn cản phản ứng polyadenine hóa (polyadenylation) của RNA.

Corepressor. Là một phân tử nhỏ kích hoạt sự kìm hãm gen trong khi giải mã bằng cách gắn chặt với một protein điều hòa.

Cosmid. Vector lai (hybrid vector) được cấu thành từ các đoạn trình tự của plasmid và các vị trí *cos* (đầu dính) của bacteriophage λ.

Cytosine (C). Base pyrimidine có trong DNA. Trên DNA mạch kép cytosine bắt cặp với một base purine là guanine (G).

D-loop. Vùng DNA ty thể nơi có một đoạn ngắn RNA bắt cặp vào để thay thế phần mạch đơn DNA nguyên thủy ở đây.

Deoxyribonucleotide triphosphate (dNTP). Tiền chất đã được triphosphoryl hóa ("năng lượng cao") cần thiết cho quá trình tổng hợp DNA. N được ký hiệu cho một trong bốn nitrogen base (A, G, T hoặc C).

Deoxyribonuclease (DNase). Loại enzyme nuclease thủy phân (phân hủy) DNA sợi đôi hoặc DNA sợi đơn.

Deoxyribonucleic acid (DNA). DNA là đại phân tử sinh học có cấu trúc xoắn đôi, tồn tại chủ yếu trong nhân tế bào, trên các nhiễm sắc thể, mang thông tin di truyền của sinh vật. Phân tử DNA gồm hai chuỗi polynucleotide, chuỗi nọ xoắn quanh chuỗi kia tạo nên chuỗi xoắn kép. Trong các nucleotide, theo chiều dọc các gốc phosphate nối xen kẽ với các phân tử đường deoxyribose tạo nên bộ khung bên ngoài của chuỗi xoắn kép, theo chiều ngang mỗi phân tử đường đều kết hợp với một trong bốn nitrogen base: adenine, guanine, cytosine hoặc thymine.

DNA không trực tiếp thể hiện chức năng sinh học mà gián tiếp qua protein do nó mã hóa. DNA tạo RNA, RNA tạo protein. RNA cũng là acid nhân (nucleic acid). Nó có

thành phần cấu tạo khá giống DNA, ngoại trừ gốc thymine (T) trong DNA được thay thế bởi gốc uracil (U), và RNA ở dạng sợi đơn chứ không phải ở dạng xoắn kép như DNA.

Quá trình đọc mã di truyền chứa trong DNA để tổng hợp protein gọi là sự phiên mã (transcription) tạo ra RNA mang thông tin di truyền là mRNA (messenger RNA). mRNA kết hợp với một cơ quan tử trong tế bào là ribosome để tạo ra protein trong quá trình dịch mã (translation). Quá trình trên được gọi là quá trình sinh học căn bản.

Năm 1962, Watson (Mỹ) và Crick (Anh) đã chia sẻ Giải Nobel với Wilkins (Anh) về phát minh ra cấu trúc không gian của DNA và ý nghĩa của nó trong việc truyền thông tin di truyền. Điều đáng tiếc là Franklin, người đã có những đóng góp đáng kể cho phát minh này đã mất trước đó. Theo qui định thì Giải Nobel không dược phép tặng cho người đã mất.

Deoxyribose. Phân tử đường có trong thành phần của DNA.

Dị polymer (heteropolymer). Một polymer bao gồm các loại monomer khác nhau. Phần lớn protein và nucleic acid là dị polymer.

Dị tương đồng (heterologous). Các trình tự gen không giống nhau hoàn toàn mà chỉ giống nhau ở một mức độ nào đó.

Dịch chuyển điểm đứt (nick translation). Phương pháp đánh dấu DNA bằng phóng xạ [α-^{32}P]dCTP nhờ enzyme DNA polymerase I của *E. coli*.

Dịch mã (translation). Sự tổng hợp protein trên khuôn mRNA. Quá trình chuyển thông tin di truyền trong trình tự base của mRNA sang trình tự amino acid của chuỗi polypeptide trong tế bào còn gọi là quá trình sinh tổng hợp protein.

Dịch mã giữ đoạn lai (hybrid-arrested translation). Kỹ thuật mà trong đó sự dịch mã của một phân tử mRNA đặc biệt (trong hệ thống dịch mã *in vitro*) bị ngăn chặn do sự hình thành đoạn lai của mRNA đó với một phân tử RNA (antisense RNA) hoặc DNA (ví dụ: cDNA) bổ sung.

Dịch mã ngược (reverse translation). Là kỹ thuật phân lập các gen nhờ khả năng của chúng trong việc lai với một đoạn mã oligonucleotide nào đó, đoạn này được chuẩn bị bằng cách dự đoán đoạn mã nucleic acid từ những đoạn mã hóa của protein biết trước.

Dideoxyribonucleotide triphosphate (ddNTP). Một đồng phân của dNTP dùng để kết thúc một chuỗi DNA trong các thí nghiệm xác định trình tự gen (sequencing).

Dimer. Là một hỗn hợp được tạo ra giữa hai phân tử đồng nhất nhưng khối lượng phân tử thì gấp đôi so với phân tử nguyên thủy.

DNA bổ sung (complementary DNA, cDNA). DNA được tổng hợp trên khuôn mẫu mRNA nhờ quá trình phiên mã ngược (reverse transcription). Do vậy, nó có trình tự sắp xếp các nucleotide bổ sung với trình tự các nucleotide trên mRNA. Ví dụ: trên mRNA trình tự đó là UUGAAG thì trên các DNA bổ sung sẽ có trình tự tương ứng là AACTTC. DNA bổ sung được tổng hợp tự nhiên trong chu trình sống của virus mang vật chất di truyền là RNA. Ví dụ: HIV, virus cúm và các retrovirus nói chung. DNA bổ sung

được tổng hợp nhân tạo trong các phòng thí nghiệm để xây dựng thư viện cDNA (cDNA library).

DNA cuộn ngược (foldback DNA). DNA có các đoạn xuôi ngược giống nhau (palindrome) hoặc các đoạn lặp lại đảo ngược (inverted repeats) nên có thể nhanh chóng gắn kết trở lại sau khi chuỗi xoắn kép DNA biến tính.

DNA khuôn mẫu (template DNA). Sợi DNA mà trình tự các nucleotide của nó được dùng làm khuôn mẫu để tổng hợp nên sợi DNA mới trong quá trình tái bản (sao chép) hoặc khuếch đại DNA (PCR) hoặc để tổng hợp nên sợi RNA mới trong quá trình phiên mã.

DNA polymerase. Enzyme tổng hợp bản sao DNA trên khuôn mẫu DNA bằng cách xúc tác phản ứng gắn từng nucleotide riêng biệt vào đầu chuỗi DNA đang được tổng hợp. Năm 1959, hai nhà khoa học người Mỹ là Kornberg và Ochoa đã được nhận Giải Nobel về những nghiên cứu đã làm sáng tỏ cơ chế cơ bản của quá trình sao chép DNA liên quan đến DNA polymerase I.

DNA replicase. Phức hợp enzyme và protein đặc hiệu, cần thiết cho sự tái bản DNA.

DNA siêu xoắn (supercoiled DNA). DNA xoắn lại trên bản thân nó, thường là kết quả của sự gấp khúc, mở xoắn hoặc xoắn lại của chuỗi xoắn kép DNA.

DNA vệ tinh (satellite DNA). Là những đoạn DNA mang các trình tự lặp lại nối tiếp có thành phần khác với trị số trung bình của DNA hệ gen. DNA vệ tinh tạo thành băng theo gradient tỷ trọng và dễ dàng phân biệt với băng của phần lớn DNA hệ gen do DNA vệ tinh có tỷ trọng nhỏ hơn. Bản sao của DNA vệ tinh lặp lại hàng triệu lần trong hệ gen, tập trung chủ yếu ở vùng tâm động và hai đầu của nhiễm sắc thể.

Dòng (clone). Tập hợp các tế bào hoặc phân tử giống hệt nhau cùng bắt nguồn từ một tế bào hay phân tử ban đầu.

Dòng phôi (germ line). Các tế bào có chức năng sinh sản, từ đó sinh ra trứng và tinh bào.

Dot blot. Là kỹ thuật trong đó các vết tròn nhỏ (hoặc các điểm) có chứa nucleic acid được đặt lên màng nitrocellulose hoặc nylon để lai với đoạn mồi DNA có đánh dấu đồng vị phóng xạ.

Đa cistron (polycistronic). Phân tử RNA mã hóa nhiều chức năng. Nhiều operon ở vi khuẩn được biểu hiện thông qua các mRNA đa cistron.

Đa hình độ dài các đoạn cắt hạn chế (restriction fragment length polymorphism, RFLP). Tính đa hình chiều dài các đoạn cắt hạn chế để chỉ các sai biệt di truyền ở vị trí nhận biết của các enzyme hạn chế (chẳng hạn như do sự thay đổi một nucleotide) dẫn đến sự sai biệt trong chiều dài của các đoạn hình thành từ phản ứng cắt

hạn chế DNA với cùng một enzyme. RFLP thường được dùng để thiết lập bản đồ di truyền với một số marker di truyền biết trước.

Đại phân tử (macromolecule). Một phân tử có khối lượng khoảng từ vài nghìn tới nhiều triệu.

Đánh dấu ở đuôi (end labelling). Bổ sung phân tử phóng xạ vào đuôi của một polynucleotide nhờ enzyme T4 polynucleotide kinase.

Đầu bằng (blunt end). Các đầu của DNA sợi đôi không có các đầu lồi 3' hoặc 5' (protruding ends).

Đầu dính (cohesive ends hoặc **sticky ends).** Các đầu của phân tử DNA có các trình tự bổ sung ngắn có thể dính kết lại để nối hai phân tử DNA với nhau. Các đầu dính thường do các enzyme hạn chế tạo ra.

Đầu tận cùng C (C terminus). Gốc carboxyl (COOH) tự do ở vị trí tận cùng của một phân tử protein hoặc chuỗi polypeptide.

Đầu tận cùng N (N terminus). Gốc amine (NH_2) ở vị trí tận cùng của một phân tử protein hoặc chuỗi polypeptide. Tất cả các polypeptide đều được tổng hợp bắt đầu từ đầu này, đến gốc carboxyl tận cùng.

Điểm đứt (nick). Điểm đứt gãy ở một sợi đơn trên DNA sợi đôi.

Điểm nóng (hot spot). Vị trí của gen có xu hướng bị đột biến nhiều hơn hẳn so với các vị trí khác trên phân tử DNA.

Điện di trên gel (gel electrophoresis). Kỹ thuật dùng để phân tách các phân tử nucleic acid hoặc protein dựa vào sự dịch chuyển của chúng trên giá thể dạng gel (agarose hoặc polyacrylamide) dưới ảnh hưởng của điện trường. Sự dịch chuyển của các phân tử này phụ thuộc vào điện tích, cấu hình, kích thước và khối lượng phân tử của nucleic acid hoặc protein cũng như dung môi và nồng độ của chất dùng làm giá thể.

Điều hòa hoạt động gen (gene operation regulation). Điều khiển sự hoạt động của các gen trong quá trình phát triển cá thể, đảm bảo cho các gen hoạt động ở thời điểm cần thiết, và do vậy quá trình phát triển cá thể diễn ra bình thường. Mô hình operon của Monod và Jacob là cơ chế điều hòa hoạt động gen đơn giản nhất.

Đính mũ (capping). Gắn nucleotide guanine (G) đã được methyl hóa (cái "mũ") vào đầu 5' của phân tử mRNA tiền thân (pre-mRNA) ở sinh vật eukaryote, cái "mũ" sau đó được giữ lại ở phân tử mRNA hoàn chỉnh (mature mRNA).

Đoạn duy nhất (unique sequence). Các đoạn DNA chỉ có một hoặc một vài bản sao trong hệ gen.

Đoạn Klenow (Klenow fragment). Còn gọi là đoạn lớn của DNA polymerase I. Đây là một đoạn của DNA polymerase I (khối lượng phân tử 76.000) của *E. coli* đã bị mất hoạt tính exonuclease 5'→3'.

Đoạn lặp lại đảo ngược (inverted repeats). Hai đoạn DNA đồng nhất, trên hai hướng ngược nhau ở cùng một phân tử. Sự lặp lại như vậy sẽ tạo ra hiện tượng palindrome.

Đoạn lặp lại cùng chiều (direct repeats). Là những chuỗi mã di truyền đồng nhất đại diện cho một hoặc hai bản sao theo cùng một hướng của một phân tử DNA.

Đoạn Okazaki (Okazaki fragment). Các đoạn DNA sợi đơn tương đối ngắn lần lượt được tổng hợp trong quá trình sao chép (tái bản) DNA theo kiểu gián đoạn dọc theo một trong hai sợi khuôn, trong khi dọc theo sợi khuôn kia thì mạch DNA mới được tổng hợp liên tục. Sau đó, các đoạn Okazaki được nối liền với nhau bằng các mối liên kết đồng hóa trị nhờ enzyme ligase.

Đoạn xếp (nested fragments). Một loạt các đoạn nucleic acid có chiều dài chỉ khác nhau một hoặc một vài nucleotide.

Đoạn xuôi ngược giống nhau (palindrome). Đoạn DNA mạch kép có trình tự sắp xếp các base trên hai mạch đơn giống hệt nhau nếu cùng được đọc theo một chiều (chẳng hạn 5'→3'). Ví dụ: các đoạn nhận biết của enzyme hạn chế.

Đóng dấu (replica plating). Phương pháp chuyển nguyên mẫu các khuẩn lạc hoặc vết tan từ một đĩa thạch gốc sang đĩa thạch mới bằng cách dùng màng nylon (ví dụ: màng Hybond-N+) vừa khít áp lên mặt thạch của đĩa gốc để dính lấy các tế bào trong các khuẩn lạc (colony) hoặc vết tan (plaque) của đĩa gốc, rồi đưa màng này áp lên mặt thạch mới.

Đơn cistron (monocistron). Mỗi mRNA monocistron chỉ mã hóa cho một protein ở eukaryote.

Đơn vị phiên mã (transcriptional unit). Đoạn DNA mã hóa cho phân tử RNA, bắt đầu từ điểm khởi đầu phiên mã đến điểm kết thúc phiên mã, nó có thể dài hơn một gen.

Đơn vị sao chép (replicon). Đoạn DNA bắt đầu từ điểm khởi đầu sao chép kéo dài về hai phía tới hai điểm kết thúc sao chép.

Đơn vị tái tổ hợp (recon). Đoạn DNA của gen có chiều dài đủ ngắn để sự trao đổi chéo không thể diễn ra ở bên trong nó được nữa. Hiện nay, được biết đó là một cặp nucleotide.

Độc tính (virulence). Khả năng của các bacteriophage gây tan tế bào vi khuẩn chủ khi chúng xâm nhiễm vào.

Đối mã (anticodon). Là một đoạn gồm ba nucleotide trên tRNA bắt cặp với cụm mã (codon) trên mRNA theo nguyên tắc bổ sung.

Đồng polymer (homopolymer). Một polymer chỉ bao gồm một loại monomer, chẳng hạn như polyphenylalanine (protein) hoặc polyadenine (nucleic acid).

Đột biến (mutation). Biến đổi trong trình tự các base của DNA. Có thể được gây nên bởi chèn đoạn, mất đoạn hoặc có các base bị sửa đổi.

Đột biến dịch khung (frameshift mutation). Đột biến làm mất hoặc thêm một cặp base ở trong gen dẫn tới hậu quả làm dịch khung dọc bình thường của mRNA đi một nucleotide. Do vậy, làm thay đổi toàn bộ thông tin di truyền đoạn gen bắt đầu từ điểm xảy ra đột biến đến cuối gen.

Đột biến đồng hoán (transition mutation). Kiểu đột biến thay thế một cặp base purine-pyrimidine này bằng một cặp base purine-pyrimidine khác tại một điểm trên DNA. Ví dụ: thay cặp GC bằng cặp AT hoặc ngược lại.

Đột biến gen (gene mutation). Còn gọi là đột biến điểm, xảy ra do những biến đổi của từng nucleotide như: mất, thêm hay thay đổi vị trí của nucleotide trên phân tử DNA. Là nguyên nhân của sự hình thành protein không bình thường. Phần lớn đột biến là có hại nhưng chúng vẫn thường được giữ lại trong quần thể vì thường là lặn và do đó có thể được duy trì trong kiểu gen mà không ảnh hưởng đến sức sống của sinh vật. Tốc độ đột biến tự nhiên thấp, nhưng tần số đột biến có thể tăng nhờ các tác nhân như phóng xạ ion hóa và các hóa chất gây đột biến gây ra.

Đột biến khuyết dưỡng (auxotrophic mutation). Còn gọi là đột biến hóa sinh. Đột biến làm mất khả năng tổng hợp một chất cần thiết cho sự sinh trưởng của tế bào.

Đột biến lỗ rò (leaky mutation). Đột biến chỉ làm giảm hoạt tính enzyme mà không làm mất hẳn hoạt tính đó.

Đột biến ngược (reverse mutation hay **reversion).** Còn gọi là hồi biến. Đột biến xảy ra theo chiều từ dạng đột biến trở về dạng bình thường, có tác dụng phục hồi trình tự nucleotide ban đầu.

Đột biến nhầm nghĩa (missense mutation). Đột biến gen khi một cặp base trong bộ ba (codon) bị thay thế bởi một base khác, dẫn đến sự thay thế amino acid ở vị trí tương ứng trên chuỗi polypeptide bằng một amino acid khác.

Đột biến ức chế (suppressor mutation). Đột biến xảy ra ở một vị trí trên DNA có tác dụng loại bỏ hoàn toàn hoặc từng phần hậu quả của đột biến ban đầu đã xảy ra ở một vị trí khác.

Đột biến vô nghĩa (nonsense mutation). Đột biến gen khi một base trên DNA bị thay thế bằng một base khác làm cho bộ ba tương ứng trên mRNA từ chỗ xác định một amino acid chuyển thành bộ ba vô nghĩa (không xác định amino acid nào), mang tín hiệu chấm dứt quá trình dịch mã. Ví dụ : amber codon (đột biến amber, amber mutation) hay ochre codon (đột biến ochre, ochre mutation). Kết quả là chuỗi polypeptide chỉ được tổng hợp đến bộ ba xảy ra đột biến vô nghĩa thì dừng lại.

Đuôi polyA (polyA tail). Đoạn trình tự dài 50-200 nucleotide adenine được bổ sung vào đầu 3' của hầu hết các mRNA eukaryote sau khi phiên mã.

***E. coli* (*Escherichia coli*).** Vi khuẩn thường có trong ruột non của động vật có xương sống. *E. coli* được coi như sinh vật mẫu cho việc nghiên cứu hoạt động của tế bào. Đây là vi khuẩn Gram âm có kích thước genome khoảng 4×10^6 base-pair. Các quá trình

biểu hiện gen (phiên mã và dịch mã) đi đôi với nhau, sinh ra sợi mRNA được tổng hợp mới và được sử dụng ngay cho quá trình dịch mã. Không có hiện tượng biến đổi sau dịch mã (post-translation). Vì thế, *E. coli* được xem là một trong những tế bào vật chủ đơn giản nhất. Rất nhiều thí nghiệm tạo dòng gen đang được thực hiện hàng ngày tại các phòng thí nghiệm đều sử dụng *E. coli* làm vật chủ với nhiều chủng khác nhau về mặt di truyền và cho những ứng dụng đặc biệt.

Endopeptidase. Enzyme xúc tác cắt liên kết peptide trong chuỗi polypeptide.

Endonuclease. Là enzyme nuclease cắt bên trong phân tử nucleic acid, ngược với exonuclease là enzyme phân giải DNA từ một đầu hoặc cả hai đầu. Nuclease thủy phân những liên kết phosphodiester giữa các nucleotide của một phân tử nucleic acid. Các nuclease có thể đặc hiệu đối với DNA (deoxyribonuclease) hoặc đặc hiệu đối với RNA (ribonuclease).

Enzyme. Chất xúc tác sinh học, là các phân tử sinh học có bản chất protein đóng vai trò chất xúc tác cho các phản ứng biến đổi hóa sinh.

Enzyme đồng phân hóa (isomerase). Enzyme xúc tác cho sự chuyển hóa một hợp chất thành một đồng phân vị trí.

Enzyme gắn DNA (DNA ligase). Enzyme dùng để nối các phân tử DNA với nhau bằng cách tạo ra mối liên kết phosphodiester giữa nhóm 5'-phosphate và nhóm 3'-hydroxyl trong quá trình tái bản hoặc sửa chữa DNA.

Enzyme hạn chế (restriction enzyme, RE). Loại endonuclease có khả năng cắt DNA tại những đoạn trình tự nhất định mà chúng nhận biết. Enzyme hạn chế được phát hiện vào năm 1970, chúng tồn tại trong tế bào vi khuẩn, có tác dụng cắt DNA ngoại lai (ví dụ: DNA của phage) tại những điểm xác định, để tiêu diệt DNA này. Cho đến nay hơn 900 enzyme hạn chế đã được tìm thấy. Các enzyme hạn chế được sử dụng rộng rãi trong các phòng thí nghiệm thao tác gen như những "chiếc kéo" cắt DNA tại những điểm đặc hiệu. Vị trí cắt phụ thuộc vào loại enzyme hạn chế được lựa chọn.

Năm 1978, Arber (Thụy Sĩ), Nathans (Mỹ) và Smith (Mỹ) đã được nhận Giải Nobel nhờ phát minh ra enzyme hạn chế và những ứng dụng của chúng để giải quyết nhiều vấn đề quan trọng của di truyền học phân tử. Các enzyme này là những "chiếc kéo phân tử" có thể cắt DNA thành những đoạn xác định, đã mở ra một thời kỳ phát triển mới của sinh học hiện đại-Thời kỳ thao tác gen.

Enzyme phiên mã ngược (reverse transcriptase). DNA polymerase phụ thuộc RNA (RNA-dependent DNA polymerase) có trong các RNA virus (retrovirus) được dùng để tổng hợp cDNA trong điều kiện *in vitro*.

Enzyme toàn phần (holoenzyme). Enzyme hoạt động gồm tất cả các tiểu đơn vị cần thiết, các nhóm ngoại và những cofactor.

Episome. Nhân tố di truyền có trong vi khuẩn, có khả năng tự sao chép trong tế bào chất, đồng thời lại có thể xen vào nhiễm sắc thể chính của vi khuẩn và sao chép đồng thời với nhiễm sắc thể.

Exon. Các đoạn DNA trong gen có chức năng phiên mã. Exon tồn tại ở cả sinh vật prokaryote và eukaryote. Riêng ở sinh vật eukaryote các exon nằm xen kẽ với các đoạn intron. Các intron chiếm tới 90% tổng số DNA của tế bào eukaryote và không có chức năng phiên mã.

Eukaryote. Sinh vật có tế bào mang nhân điển hình (nhân thật) nghĩa là nhân được bao bọc bởi màng nhân và tham gia vào hai cơ chế phân bào quan trọng là nguyên phân và giảm phân.

Exonuclease. Loại enzyme nuclease chỉ tác động vào đầu tận cùng của phân tử nucleic acid, cắt ra từng nucleotide một theo thời gian. Chúng có thể chuyển hóa theo đầu 5' hoặc 3' của sợi DNA.

***Ex vivo*.** Thuật ngữ dùng để chỉ các thí nghiệm thực hiện trên tế bào nuôi cấy, các tế bào này sau đó sẽ được đưa vào một cơ thể sống.

S-galactosidase. Enzyme được mã hóa bởi gen *lacZ*. Enzyme này thủy phân lactose thành glucose và galactose.

Gây đột biến định hướng bằng oligonucleotide (oligonucleotide-directed mutagenesis). Còn gọi là gây đột biến định hướng điểm (site-directed mutagenesis), là quá trình tạo ra một biến đổi xác định trên DNA bằng cách sử dụng một oligonucleotide tổng hợp (primer) có một vài biến đổi trong trình tự nucleotide.

Gel chậm (gel retardation). Phương pháp xác định điểm bám của protein trên các đoạn DNA, dựa vào độ di chuyển chậm của chúng so với DNA không bị protein bám trong các thí nghiệm điện di trên gel.

Gen (gene). Là đơn vị di truyền, yếu tố quyết định một tính trạng cơ thể. Thông tin di truyền của các gen được mã hóa trong DNA quyết định tính biến dị của loài và của cá thể. DNA là một chuỗi bao gồm các đơn vị nucleotide, có bốn loại nucleotide mang bốn nitrogen base khác nhau là adenine (A), guanine (G), cytosine (C), và thymine (T). Trình tự các nucleotide của một gen xác định một polypeptide hoặc một RNA. Gen có khả năng bị đột biến. Các gen chủ yếu nằm dọc theo nhiễm sắc thể ở trong nhân tế bào. Mỗi gen chiếm một vị trí xác định trên nhiễm sắc thể gọi là locus. Gen có thể tồn tại ở nhiều dạng gọi là các allele. Các gen biểu hiện thông qua các phân tử do chúng sinh ra là RNA (trong quá trình phiên mã) và protein (trong quá trình dịch mã).

Gen cấu trúc (structural gene) hay **gen mã hóa (coding gene).** Là trình tự mã hóa sản phẩm protein.

Gen chỉ thị (reporter gene). Là một gen mã hóa mà sản phẩm của nó được trắc nghiệm một cách dễ dàng (ví dụ chloramphenicol acetyltranferase), nó có thể được gắn

với bất cứ một promoter nào sao cho sự biểu hiện của gen này được dùng để thử nghiệm chức năng của promoter.

Gen điều hòa (regulatory gene). Một gen mà sản phẩm protein của nó tham gia vào sự điều hòa biểu hiện của một gen khác. Ví dụ: gen mã hóa một protein kìm hãm.

Gen gây chết (death gene). Gen làm chết cá thể ở một giai đoạn nào đó trong quá trình phát triển của chúng.

Gen nhảy (transposon). Là một đoạn DNA nhờ có cấu trúc đặc biệt nên có khả năng di chuyển từ một vị trí này đến một vị trí khác trên phân tử DNA hay từ một nhiễm sắc thể này sang một nhiễm sắc thể khác trong tế bào lưỡng bội. Gen nhảy chỉ chuyển đến những vị trí xác định trong hệ gen. Trong nhiều trường hợp gen nhảy có thể gây đột biến tại vị trí nó di chuyển đến. Gen nhảy có thể dùng làm phương tiện để gây đột biến định hướng. Thường các gen này mang tính kháng kháng sinh giúp vi khuẩn thích nghi nhanh với thuốc kháng sinh.

Năm 1983, McClintock (Mỹ) đã nhận Giải Nobel nhờ khám phá ra các nhân tố di truyền vận động (gen nhảy) trên đối tượng cây ngô.

Gen ức chế (repressor gene). Có hai chức năng: (1) Là một gen điều hòa (regulator) mà sản phẩm của nó là một protein có tác dụng ức chế hoạt động phiên mã của một operon. (2) Là một gen phục hồi (suppressor) kiểu hình ban đầu của đột biến trong một gen khác.

Ghép đôi lệch (mismatch). Sự ghép đôi không đúng với quy luật bổ sung giữa các nucleotide thuộc hai sợi đơn DNA trong mạch kép.

Ghép exon hay **splicing (RNA).** Quá trình cắt bỏ những intron và nối các exon của sản phẩm phiên mã ban đầu (tiền thân mRNA) để tạo thành mRNA hoàn chỉnh (mature mRNA). Quá trình biến đổi này xảy ra trong nhân tế bào.

Giả thuyết một gen-một enzyme (one gene-one enzyme hypothesis). Giả thuyết dựa trên các công trình của Beadle và Tatum về di truyền học hóa sinh, theo đó mỗi gen kiểm soát sự tổng hợp một enzyme.

Giảm phân (meiosis). Sự phân bào trải qua hai giai đoạn của một tế bào lưỡng bội (2n) dẫn đến sự hình thành các giao tử đơn bội (1n) hoặc các bào tử hữu tính mang một nửa vật chất di truyền của tế bào lưỡng bội ban đầu.

Giảm phân I (meiosis I). Lần phân chia thứ nhất của giảm phân làm giảm số lượng nhiễm sắc thể đi một nửa và tạo ra các tổ hợp nhiễm sắc thể mới. Lần phân chia này gồm bốn giai đoạn: pha đầu, pha giữa, pha sau và pha cuối.

Giảm phân II (meiosis II). Lần phân chia thứ hai của giảm phân dẫn đến sự phân tách các nhiễm sắc tử và hoàn thành quá trình giảm phân nói chung.

Gốc tái bản (origin, *ori*). Trình tự nucleotide hoặc vị trí trên DNA mà ở đó bắt đầu sự tái bản (sao chép).

Guanine (G). Một trong năm nitrogen base có trong thành phần của DNA và RNA, bắt cặp với cytosine (C) trong chuỗi xoắn kép nhờ ba liên kết hydrogen.

Helicase. Enzyme xúc tác việc tháo xoắn của chuỗi xoắn kép DNA trong quá trình sao chép ở *E. coli*.

Hệ gen hay **bộ gen (genome).** Là tập hợp các gen có trong một tế bào đơn bội eukaryote, trong một tế bào prokaryote hoặc trong một virus. Hệ gen chứa toàn bộ DNA của cơ thể, ví dụ: hệ gen người chứa DNA dài khoảng 1,6 m nhưng chỉ rộng khoảng 5 phần tỉ mm. Ở người, số DNA nói trên được chia làm 46 phần có độ dài ngắn khác nhau gọi là nhiễm sắc thể (chromosome), là tập hợp DNA ở dạng nén chặt đến kích thước đường kính chỉ còn 3-4 phần triệu mét. Tuy nhỏ như vậy, nhưng nhiễm sắc thể người có đến 3 tỷ gốc nucleotide. Sự sắp xếp đặc thù của chúng quyết định bản chất sinh học của cơ thể.

Hiệu ứng glucose (glucose effect). Hiện tượng khử hoạt tính của operon cảm ứng ở vi khuẩn khi có mặt glucose kể cả khi hiện diện chất cảm ứng operon.

Histone. Nhóm protein kiềm nằm trong phức hợp với DNA ở các nhiễm sắc thể eukaryote và đóng vai trò quan trọng trong việc xác định cấu trúc của nhiễm sắc thể.

Hoán chuyển gen (gene conversion). Sự hoán chuyển gen là sự thay đổi một mạch đơn của chuỗi xoắn kép DNA để giúp nó bổ sung bằng một mạch đơn khác ở bất cứ vị trí nào mà nó đã bị mất cặp nối base.

Hoạt tính đặc hiệu (specific activity). Là hoạt độ phóng xạ trên một đơn vị nguyên liệu, chẳng hạn: một mẫu dò (probe) đánh dấu phóng xạ có thể có hoạt tính đặc hiệu 106 lần đếm/phút trên microgram. Hoạt tính đặc hiệu cũng được dùng để xác định hoạt độ của enzyme.

Hồi tính (renaturation). Sự tái bắt cặp của hai mạch đơn DNA bổ sung của một phân tử DNA mạch kép.

Hộp CAAT (CAAT box). Đoạn DNA nằm ngược hướng cách điểm bắt đầu phiên mã khoảng 75 nucleotide hiện diện ở sinh vật eukaryote. Đây là một trong những đoạn tăng cường sự bám của RNA polymerase.

Hộp Goldberg-Hogness (Golderg-Hogness box) hay **hộp Hogness (Hogness box)** hay **hộp TATA (TATA box).** Đoạn DNA có trong các promoter eukaryote tương tự hộp Pribnow có trong các sinh vật prokaryote, nằm cách vị trí bắt đầu phiên mã khoảng 30 nucleotide, được coi là operator ở sinh vật eukaryote, có trình tự phổ biến của sợi bên là TATAAAT.

Hộp Pribnow (Pribnow box). Một phần của promoter trong các hệ gen prokaryote, nằm cách phía trên điểm khởi đầu phiên mã 10 nucleotide. Trình tự của sợi tương đồng với hộp Pribnow là TATAAT.

In dấu DNA (DNA fingerprinting) hay in dấu di truyền (genetic fingerprinting). Là phương pháp dùng các mẫu dò phóng xạ hoặc dùng kỹ thuật PCR để nhận dạng các băng DNA có các đoạn lặp lại với tần số cao. Bản mẫu hình các băng DNA là duy nhất đối với mỗi cá thể, và do vậy có thể dùng để xác định đặc trưng cá thể hoặc quan hệ huyết thống.

In dấu chân DNA (DNA footprinting). Phương pháp nhận dạng các vùng DNA mà các protein điều hòa bám vào.

Intron. Những đoạn DNA nhỏ ở sinh vật eukaryote không mang thông tin mã hóa amino acid, phân bố rải rác dọc theo phân tử DNA. Sau khi thông tin từ DNA được phiên mã sang mRNA thì các intron trên mRNA bị cắt bỏ, các đoạn mRNA còn lại gồm toàn các exon được nối lại với nhau và chuyển đến ribosome để dùng làm khuôn mẫu cho quá trình dịch mã. Intron không thấy có ở sinh vật prokaryote.

In vitro và ***in vivo*.** Là thuật ngữ mô tả thí nghiệm trong ống nghiệm (*in vitro*) và trong cơ thể (*in vivo*). Cùng với sự phát triển ứng dụng của máy tính, hiện nay các nhà khoa học còn tiến hành thí nghiệm mô phỏng bằng computer. Quá trình này gọi là thí nghiệm *in silico*.

Isozymes hay **isoenzymes.** Nhóm các enzyme cùng xúc tác một phản ứng nhưng khác nhau về cấu trúc sơ cấp hoặc tốc độ điện di, thuật ngữ allozymes cũng có ý nghĩa tương tự.

Kéo dài đoạn mồi (primer extension). Sự tổng hợp một bản sao nucleic acid bắt đầu từ đoạn mồi. Được sử dụng để đánh dấu phóng xạ đoạn DNA làm mẫu dò hoặc khuếch đại một đoạn DNA bằng kỹ thuật PCR.

Khoảng cách bản đồ (map distance). Số trao đổi chéo trung bình xảy ra giữa hai locus gen trên nhiễm sắc thể hoặc trên sợi DNA.

Khuếch đại (amplification). Sự sản xuất nhiều bản sao của một trình tự DNA.

Khung đọc mở (open reading frame, ORF). Là một trình tự mã hóa chuỗi polypeptide được bắt đầu với mã khởi đầu (initiator codon) và kết thúc bằng một mã dừng (stop codon). Một khung đọc mở bị ngăn chặn nếu một stop codon được định vị gần với mã khởi đầu. Mặc dù về lý thuyết mã di truyền được xây dựng dựa trên bộ ba nucleotide do đó sẽ có ba khung đọc có thể có trên mỗi sợi DNA, tuy nhiên trong thực tế khung đọc chính xác được xác định bởi một điểm bắt đầu cố định.

Khuyết đoạn (deletion, deficiency). Đột biến nhiễm sắc thể dẫn đến làm mất một đoạn vật chất di truyền và thông tin di truyền chứa trong nó rời khỏi nhiễm sắc thể.

Kiểu gen (genotype). Cấu trúc di truyền của cơ thể. Hiện trạng của cá thể (kiểu hình) phụ thuộc vào quan hệ trội-lặn giữa các allele trong kiểu gen và vào mối tương tác giữa kiểu gen và môi trường.

Kiểu hình (phenotype). Ngoại hình của một sinh vật, là kết quả tương tác giữa kiểu gen với môi trường. Nhiều gen có trong kiểu gen nhưng không biểu hiện ra kiểu

hình vì chúng bị che khuất bởi các allele trội. Những sinh vật có cùng kiểu gen có thể có các kiểu hình rất khác nhau trong những môi trường khác nhau. Kiểu hình của cá thể đồng hợp tử về allele lặn gọi là kiểu hình lặn.

Kiểu hoang dại (wild type). Dạng thường thấy nhất của một gen trong quần thể hoang dại. Allele kiểu hoang dại được ký hiệu bằng một chữ in hoa hoặc thêm dấu cộng sau chữ viết thường, ví dụ: A hay a^+. Allele kiểu hoang dại thường là trội và cho kiểu hình bình thường.

Kiểu nhân (karyotype). Bộ nhiễm sắc thể đặc trưng của một loài với số lượng, kích thước và hình thái xác định của các nhiễm sắc thể. Mỗi loài được đặc trưng bởi kiểu nhân riêng của nó. Các loài khác nhau có kiểu nhân khác nhau.

Kilobase (kb). 1000 base (hoặc cặp base), được dùng như đơn vị để đo hoặc xác định chiều dài của các phân tử DNA hoặc RNA.

Kinase. Các enzyme xúc tác phản ứng phosphoryl hóa một phân tử nhận nhờ ATP.

Lai (hybridization). Ghép các phân tử nucleic acid tách biệt lại với nhau thông qua các mối liên kết hydrogen giữa các base bổ sung.

Lai khuẩn lạc (colony hybridization). Kỹ thuật lai *in situ* để xác định vi khuẩn mang vector khảm (chimeric vector) mà DNA của vector này tương đồng với một đoạn mã hóa đặc biệt nào đó.

Lai phân tử (molecular hybridization). Quá trình trong đó hai mạch nucleic acid bổ sung (A-T, G-C) bắt cặp hình thành nên một mạch kép. Đây là một kỹ thuật hữu ích để phát hiện một trình tự nucleotide chuyên biệt.

Lai tại chỗ (*in situ* hybridization). Quá trình bắt cặp giữa mẫu dò (là một trình tự DNA sợi đơn hay RNA) với DNA của tế bào được cố định trên lam kính.

Lập bản đồ hạn chế (restriction mapping). Kỹ thuật dùng để xác định vị trí các điểm cắt hạn chế trên phân tử DNA.

Liên kết peptide (peptide bond). Mối liên kết đồng hóa trị trong chuỗi polypeptide, nối nhóm α-carboxyl của một amino acid với nhóm α-amine của một amino acid kế tiếp.

Liên kết phosphodiester (phosphodiester bond). Một phân tử chứa hai alcol, ester hóa với một phân tử phosphoric acid làm cầu nối giữa chúng với nhau. Ví dụ: chuỗi polynucleotide được nối bởi liên kết 5'→3' phosphodiester giữa hai nucleotide cạnh nhau.

Linker. Một oligonucleotide tổng hợp có hai đầu bằng, chứa một hoặc nhiều vị trí cắt hạn chế cho phép nối cDNA sợi đôi với các plasmid vector hoặc bacteriophage λ vector. cDNA sợi đôi trước đó được xử lý với DNA polymerase của bacteriophage T4 hoặc DNA polymerase I của *E. coli* để tạo đầu bằng. Các linker sau khi gắn với hai đầu bằng của đoạn cDNA nhờ DNA ligase sẽ được cắt hạn chế để tạo ra đầu so le tương đồng

với hai đầu của vector. Phản ứng gắn giữa đoạn cDNA có mang linker ở hai đầu với vector cũng được xúc tác nhờ DNA ligase.

Locus. Vị trí mà gen định vị trên nhiễm sắc thể.

LTR (long-terminated repeat). Là một đoạn mã được lặp lại trực tiếp ở cả hai đầu của một retroviral DNA.

Lysosome. Một bào quan có màng bao bọc ở trong tế bào chất của những tế bào eukaryote. Lysosome chứa nhiều enzyme thủy phân.

Lý thuyết trung tâm (central dogma). Thuyết cho rằng thông tin được truyền theo một chiều từ DNA đến RNA và đến protein.

Mapping and sequencing gene. Lập bản đồ gen và xác định trình tự các nucleotide.

Mã di truyền (codon). Nhóm ba nucleotide nằm kề nhau (bộ ba) trên phân tử mRNA xác định một amino acid trên chuỗi polypeptide, hoặc là tín hiệu kết thúc việc tổng hợp polypeptide.

Mã thoái biến (degenerate codon). Mã di truyền mà ở đó một amino acid được quy định bởi một số bộ ba nitrogen base, chứ không phải chỉ bởi một bộ ba. Thoái biến là đặc điểm vốn có của mã di truyền tồn tại phổ biến ở sinh giới.

Mã vô nghĩa (nonsense codon) hay **mã dừng (stop codon).** Là codon mà ở đó quá trình dịch mã dừng lại (nơi kết thúc của chuỗi polypeptide). Có tất cả ba codon vô nghĩa với các tên gọi là amber (UAG), ocher (UAA) và opal (UGA)

Maturation. Quá trình trong đó các mRNA vừa được phiên mã trải qua một số biến đổi hóa học để trở thành mRNA hoàn chỉnh sẵn sàng làm khuôn mẫu cho việc tổng hợp protein.

Máy đếm nhấp nháy (scintillation counter). Máy dùng để xác định hoạt tính phóng xạ trong một mẫu thí nghiệm.

Mẫu dò DNA (DNA probe). Một đoạn RNA hay DNA chuyên biệt được đánh dấu bằng đồng vị phóng xạ hay bằng hóa chất (chất phát huỳnh quang hoặc enzyme), dùng để định vị một trình tự nucleic acid nhất định thông qua kỹ thuật lai phân tử (xem Northern blot, Southern blot...)

Monomer. Là các phân tử đơn vị nhỏ, có thể liên kết với các phân tử đơn vị giống nó để hình thành những phân tử lớn hơn (polymer). Ví dụ: các nucleotide là các monomer của nucleic acid và các amino acid là monomer của protein.

Mô hình bảo toàn. Mô hình sao chép DNA mà theo đó hai sợi DNA bố mẹ vẫn được giữ nguyên và làm khuôn để tổng hợp chuỗi xoắn kép mới (chuỗi con).

Mồi (primer). Một trình tự DNA hay RNA ngắn, bắt cặp với một mạch của DNA khuôn mẫu và có mang một đầu 3'-OH tự do giúp DNA polymerase bắt đầu tổng hợp một chuỗi DNA mới.

Mũ chụp (cap). Một nucleotide biến đổi, 7-methyl guanine, gắn vào đầu 5' của base đầu tiên ở mọi chuỗi mRNA ngay khi bắt đầu phiên mã. Mũ cần thiết cho sự hoàn chỉnh, sự ổn định và sự dịch mã của mRNA.

Ngoại tế bào (exocytosis). Hiện tượng phóng thích vật chất nội bào ra ngoài tế bào.

Nguyên dưỡng (prototroph). Các tế bào có thể mọc trên môi trường không cần bổ sung thêm chất dinh dưỡng.

Nguyên phân (mitosis). Tái bản các nhiễm sắc thể trong tế bào sinh dưỡng eukaryote. Sự phân chia những tế bào eukaryote kéo theo sự phân chia nhân gọi là mitosis. Quá trình phân chia tế bào qua năm giai đoạn (prophase, prometaphase, metaphase, anaphase và telophase), sau đó một tế bào được phân chia và tạo ra hai tế bào có tính chất di truyền giống với tế bào mẹ.

Nhân tố σ (housekeeping σ factor). Là một nhân tố khởi đầu có bản chất enzyme của sinh vật prokaryote, nó là một phần của RNA polymerase để gắn với các vị trí promoter trên DNA trong quá trình phiên mã. Các nhân tố σ khác nhau được hoạt hóa trong phản ứng với các điều kiện môi trường khác nhau. Mỗi phân tử RNA polymerase có chính xác một nhân tố σ tiểu đơn vị. Vi khuẩn *E. coli* có bảy nhân tố σ, trong khi các loại vi khuẩn khác có số lượng các nhân tố σ rất khác nhau.

Nhân tố giới tính (fertility factor, F). Là nhân tố di truyền ở vi khuẩn xác định giới tính "đực" hoặc "cái" của chúng. Ở *E. coli*, nhân tố hữu thụ ký hiệu là F xác định giới tính "đực". Khi trong tế bào không có nhân tố F thì đó là tế bào "cái", ký hiệu là F^-. Khi nhân tố F đính vào nhiễm sắc thể thì đó là tế bào "siêu đực" ký hiệu là Hfr.

Nhân tố kéo dài (elongation factor, EF). Protein đặc hiệu tham gia vào việc đưa các phân tử aminoacyl-tRNA đến ribosome để cho amino acid trong phức hệ phân tử nói trên có thể gắn vào chuỗi polypeptide đang được tổng hợp.

Nhân tố kết thúc (termination factor). Yếu tố protein trong cytosol cần thiết cho việc giải phóng chuỗi polypeptide đã hoàn thành ở trên ribosome (cũng gọi là release factors-RF).

Nhân tố khởi đầu (initiation factor). Ở prokaryote viết tắt là IF, còn ở eukaryote là eIF. Là các protein kết hợp với tiểu đơn vị của ribosome vào giai đoạn khởi động của quá trình dịch mã.

Nhân tố Rho (Rho factor). Một protein có vai trò kết thúc sự tổng hợp một số mRNA.

Nhiễm sắc thể (chromosome). Một thành phần của tế bào, bao gồm các protein và một phân tử DNA độc nhất. DNA chứa nhiều gen, có chức năng dự trữ và truyền thông tin di truyền. Số lượng nhiễm sắc thể trong một tế bào có thể thay đổi từ 1 (ở vi khuẩn) cho tới trên 100, nhưng không thay đổi ở một loài nhất định (ví dụ: 46 đối với người).

Nhiễm sắc thể thường (autosome). Tất cả nhiễm sắc thể trừ những nhiễm sắc thể giới tính hoặc nhiễm sắc thể của ty thể.

Nhiễm sắc tử (chromatide). Hai chuỗi song song và giống nhau, nối với nhau bởi phần tâm của một nhiễm sắc thể dạng đơn đã được nhân đôi, thực thể này tồn tại cho đến khi bắt đầu kỳ sau (anaphase).

Northern blot. Kỹ thuật chuyển và cố định RNA từ formaldehyde agarose gel (sau khi được phân đoạn bằng điện di) lên màng lai bằng nylon hoặc nitrocellulose để lai với mẫu dò (probe) được đánh dấu đồng vị phóng xạ [α-^{32}P]dCTP hoặc digoxigenin-dUTP. Tín hiệu lai sau đó được phát hiện trên phim X-quang (trường hợp [α-^{32}P]dCTP) hoặc trên màng lai (trường hợp digoxigenin-dUTP).

Nội thực bào (endocytosis). Sự xâm nhập của chất ngoài tế bào vào trong tế bào.

Nuclease S1. Enzyme thủy phân DNA sợi đơn.

Nucleic acids. Những polynucleotide sinh học thiên nhiên, trong đó những đơn vị nucleotide được kết hợp với nhau bởi những liên kết phosphodieste thành trình tự DNA hoặc RNA riêng biệt.

Nucleolar organizer. Là vùng mang gen mã hóa cho rRNA của một nhiễm sắc thể.

Nucleolus. Là một vùng đặc biệt của nhân, được tạo ra do sự giải mã của những gen rRNA.

Nucleoside. Một hợp chất gồm một base purine hoặc pyrimidine kết hợp đồng hóa trị với một phân tử đường pentose.

Nucleosome. Đơn vị cấu trúc của nhiễm sắc chất, gồm một đoạn DNA có kích thước khoảng 200 bp quấn bao quanh một lõi tập hợp 8 protein histone.

Nucleotide. Một nucleoside phosphoryl hóa với một trong những hydroxyl của pentose. Phân tử đóng vai trò cấu trúc cơ sở của DNA và RNA, gồm ba phần: đường pentose (ribose trong RNA, deoxyribose trong DNA), nitrogen base và gốc phosphate.

Nucleolytic. Phản ứng thủy phân một cầu nối phosphodiester trong một nucleic acid.

Nuclease Bal 31. Một loại enzyme exonuclease thủy phân cả hai sợi của phân tử DNA cùng một lúc.

Oligo. Tiếp đầu ngữ có nghĩa là "một ít", ví dụ: oligonucleotide (polynucleotide) có một ít nucleotide hoặc oligopeptide (polypeptide) có một ít peptide.

Oligo(dT)-cellulose. Một đoạn ngắn gồm các gốc deoxythymidine liên kết với cơ chất cellulose, được sử dụng để tinh sạch mRNA eukaryote.

Oligomer. Thuật ngữ chung để chỉ một đoạn ngắn các monomer.

Oligonucleotide. Một đoạn ngắn các monomer là nucleotide, thường từ 20-30 nucleotide.

Operon. Nhóm các gen ở vi khuẩn chịu sự điều khiển chung của một gen điều hòa. Cấu trúc operon bao gồm: 1) Nhóm gen cấu trúc mà hoạt động của chúng chịu sự điều khiển chung, 2) Gen điều hòa sản sinh ra protein điều hòa, 3) Vùng chỉ huy (operator) và

trình tự khởi động (promoter) chịu tác động của protein điều hòa. Protein điều hòa có tác dụng bám vào vùng chỉ huy làm cho nó bị "đóng", do vậy nhóm gen cấu trúc ngừng hoạt động. Khi trong môi trường xuất hiện cơ chất của các gen cấu trúc thì cơ chất này bất hoạt hóa protein điều hòa khiến nó không bám vào vùng chỉ huy được nữa, vùng chỉ huy được "mở" và nhóm gen cấu trúc lại hoạt động.

Operon chịu ức chế (repressible operons). Các operon sinh tổng hợp chuỗi amino acid bị ức chế khi một chất (ví dụ: chuỗi amino acid đó) được đưa vào môi trường.

Orche codon. Chuỗi kết thúc bằng codon UAA.

Opal codon. Chuỗi kết thúc bằng codon UGA.

Ôn hòa (temperate). Trạng thái tiềm tan của các bacteriophage tế bào vật chủ.

ɼ-peptide. Một phần của protein β-galactosidase, được mã hóa bởi đoạn gen *lacZ*.

Periplasm. Phần ngoại bào tương nằm giữa lớp màng ngoài và màng trong của vi khuẩn.

Peroxisome. Bào quan có trong tế bào chất eukaryote chứa enzyme xúc tác cho sự tạo thành và phân hủy peroxide (H_2O_2).

Phage. Viết tắt của bacteriophage (thực khuẩn thể), là loại virus xâm nhiễm và sinh sản bên trong vi khuẩn. Phage thường có vỏ bọc protein, phức hợp bao gồm phần đầu có hình đa diện chứa nucleic acid (DNA hoặc RNA) và đuôi mà qua đó nucleic acid xâm nhập vào vi khuẩn chủ. Sau quá trình nhân lên của nucleic acid của phage, tế bào vi khuẩn chủ thường bị tan biến. Loại phage luôn luôn làm tan tế bào vi khuẩn khi chúng xâm nhiễm vi khuẩn gọi là phage độc. Ví dụ: phage T4. Ngược lại, còn có phage ôn hòa, khi xâm nhiễm vi khuẩn nó gây nên phản ứng tiềm tan, nghĩa là hệ gen của phage gắn vào nhiễm sắc thể vi khuẩn và được sao chép cùng với nhiễm sắc thể đó. Hệ gen của phage ở trạng thái gắn như vậy với nhiễm sắc thể vi khuẩn gọi là prophage.

Phagemid. Là một loại plasmid vector có mang các đoạn trình tự của phage.

Phản ứng chuỗi polymerase (polymerase chain reaction, PCR). Phương pháp dùng trong phòng thí nghiệm để khuếch đại các đoạn DNA đặc biệt lên hàng triệu lần trong vòng vài giờ thông qua 20-30 chu kỳ nhiệt, mỗi chu kỳ bao gồm ba mức nhiệt độ: biến tính ở 90-95^{o}C, bắt cặp với mồi ở 40-65^{o}C hoặc hơn và tổng hợp mạch mới nhờ DNA polymerase chịu nhiệt (Taq polymerase) ở 70-72^{o}C. PCR có ứng dụng rộng rãi trong chẩn đoán y học, phân tích sự đa dạng sinh học, chọn giống và trong nhiều lĩnh vực khác.

Nhà khoa học Mỹ (Tiến sĩ Mullis) người phát minh ra kỹ thuật PCR đã nhận giải Nobel năm 1993. Cùng chia sẻ Giải Nobel với Mullis là Smith (Canada) do có những đóng góp mang tính nền tảng cho việc gây đột biến điểm định hướng, dựa trên các oligonucleotide và việc phát triển chúng trong các nghiên cứu protein.

Phản ứng tập trung (condensation reaction). Là phản ứng trong đó cầu nối được hình thành với sự mất đi một phân tử nước và thêm vào một amino acid trong quá trình hình thành chuỗi polypeptide.

Phản ứng thủy phân (hydrolysis). Phản ứng trong đó một liên kết cộng hóa trị bị phá vỡ khi có sự kết hợp một phân tử nước.

Phát sinh đột biến (mutagenesis). Quá trình tạo ra đột biến trong DNA.

Phân tích trình tự gen (gene sequencing). Là kỹ thuật xác định trình tự theo cấu trúc bậc một của chuỗi các nucleotide trong một phân tử nucleic acid. Phân tích trình tự của DNA có các phương pháp hóa học của Maxam-Gilbert và phương pháp enzyme của Sanger. Trong những năm gần đây, một số phương pháp xác định trình tự mới nhờ sự hỗ trợ của máy tính đã xuất hiện. Bên cạnh kỹ thuật thông thường sử dụng các polyacrylamide gel để phân ly các phân tử DNA có độ dài khác nhau, các kỹ thuật mới liên quan đến phát hiện huỳnh quang của các nucleotide được đánh dấu, phân tích trình tự DNA bằng khối phổ, điện di mao dẫn hoặc lai với các đoạn oligonucleotide được tổng hợp nhân tạo cũng đã ra đời. Năm 1980, Sanger (Anh) và Gilbert (Mỹ) đã được trao giải Nobel do đã có những đóng góp quan trọn về phương pháp xác định trình tự các nucleotide trong phân tử DNA. Đóng góp này là mốc lịch sử to lớn trong sinh học phân tử, là nguyên lý của tất cả các máy xác định trình tự DNA tự động đang sử dụng hiện nay trên khắp thế giới.

Phần cuối (telomere). Đoạn cuối, phần cuối của một nhiễm sắc thể thẳng của eukaryote, bao gồm những trình tự DNA ngắn được lặp lại nhiều lần.

Phần tâm (centromere). Phần co thắt được thấy trên nhiễm sắc thể ở kỳ giữa, đó là nơi hai nhiễm sắc tử đính với nhau.

Phiên mã (transcription). Quá trình tổng hợp mRNA từ khuôn mẫu DNA.

Phiên mã ngược (reverse transcription). Quá trình tổng hợp DNA từ khuôn mẫu mRNA nhờ enzyme phiên mã ngược (reverse transcriptase).

Phóng xạ tự ghi (autoradiography). Kỹ thuật phát hiện các phân tử có đánh dấu phóng xạ thông qua hiệu ứng tạo ảnh của các phân tử này trên phim X-quang.

Phosphoryl hóa (phosphorylation). Phản ứng tạo thành một dẫn xuất phosphate của một phân tử sinh học dưới tác dụng xúc tác của một enzyme.

Phosphatase kiềm (alkaline phosphatase). Enzyme loại bỏ các nhóm 5'-phosphate từ đầu của các phân tử DNA và để lại các nhóm 5'-hydroxyl.

Phối tử (ligand). Một phân tử hoặc một ion kết hợp với một protein, ví dụ: hormone có thể kết hợp với thụ thể (receptor) đặc hiệu của nó.

Plasmid. Là DNA vi khuẩn có cấu trúc mạch vòng kép, nằm trong tế bào chất và ngoài nhân, có khả năng sao chép độc lập đối với nhiễm sắc thể của tế bào. Tồn tại cả ở sinh vật prokaryote và eukaryote. Ngày nay, các plasmid thiết kế nhân tạo được sử dụng rộng rãi như là các vector dùng trong các kỹ thuật tạo dòng và biểu hiện gen.

Polynucleotide. Trình tự những nucleotide nối đồng hóa trị với nhau, trong đó vị trí 3' của pentose của một trong những nucleotide được nối với một liên kết phosphodieste ở vị trí 5' của pentose của nucleotide tiếp theo.

Polypeptide. Một chuỗi dài những amino acid nối với nhau bởi những liên kết peptide.

Polysome (polyribosome). Cấu trúc gồm nhiều ribosome kết hợp với mRNA trong quá trình dịch mã. Những polysome tự do trong tế bào chất, tổng hợp những protein nội bào. Những polysome kết hợp với màng ngoài của lưới nội sinh chất, tổng hợp những protein dùng cho sự tổng hợp màng hoặc được bài tiết ra ngoài.

Primase (protein Dna G). Là một dạng của RNA polymerase và là sản phẩm của gen Dna G. Trong vi khuẩn, primase liên kết với DNA helicase tạo thành một phúc hợp gọi là primosome. Primase được hoạt hóa bởi DNA helicase nhờ đó nó sẽ tổng hợp một RNA primer ngắn xấp xỉ 11±1 nucleotide, enzyme DNA polymerase sẽ kéo dài RNA primer bằng cách bổ sung các nucleotide mới. Vì thế, primase là một nhân tố quan trọng trong tái bản DNA bởi vì nếu không có nó DNA polymerase không thể khởi đầu sự tổng hợp một sợi DNA mới.

Prokaryote. Sinh vật đơn bào không có nhân tế bào điển hình, DNA nằm trong tế bào chất không có màng bao bọc, không có nguyên phân và giảm phân; đại diện điển hình là vi khuẩn.

Prophage. Phage ôn hòa đã xen vào nhiễm sắc thể của vi khuẩn tiềm tan. Nó sao chép đồng thời với nhiễm sắc thể của tế bào vi khuẩn chủ.

Protein. Một phân tử lớn gồm một hoặc nhiều chuỗi polypeptide, mỗi chuỗi có một trình tự amino acid và một khối lượng phân tử đặc trưng. Protein là hợp chất quan trọng bậc nhất đối với cơ thể sống. Về cấu trúc, protein là phân tử mạch dài gồm các đơn vị cấu trúc nhỏ là các amino acid nối với nhau qua mối liên kết peptide. Khối lượng phân tử của protein từ vài nghìn đến vài triệu. Có khoảng 20 loại amino acid. Các loại protein phức tạp hơn có liên kết thêm với các nhóm bổ sung.

Protein dung hợp (fusion protein). Là một protein tái tổ hợp lai được mã hóa bởi một gen lai (fusion gene) do sự dung hợp *in vitro* các đoạn gen khác nhau trên plasmid vector và sau đó biến nạp vào vi sinh vật chủ (chẳng hạn *E. coli*). Vì vậy, protein dung hợp sẽ mang trình tự amino acid của hai protein khác biệt được tổng hợp từ đầu N của vector biểu hiện.

Protein không histone (non-histone). Protein acid liên kết với DNA trong các nhiễm sắc thể eukaryote. Còn gọi là protein không kiềm.

Protein nguyên thể (native protein). Là một protein tái tổ hợp được mã hóa bởi một gen ngoại lai (foreign gene) trong vi sinh vật chủ. Khác với protein dung hợp, protein nguyên thể được tổng hợp từ đầu N của nó chứ không phải từ đầu N của vector.

Protein sợi (fibrous protein). Những protein không tan, có chức năng bảo vệ hoặc cấu trúc, chứa những chuỗi polypeptide có cấu trúc bậc hai.

Purine. Một hợp chất dị vòng, kiềm, có nitrogen, là thành phần của những nucleotide và nucleic acid. Purine chứa một nhân pyrimidine kết hợp với một nhân imidazol.

Purmycin. Một loại kháng sinh có tác dụng ức chế sự tổng hợp polypeptide bằng cách gắn vào chuỗi polypeptide, cạnh tranh với aminoacyl-tRNA gắn lên vị trí của ribosome, và như vậy tạo ra sự kết thúc sớm tổng hợp chuỗi polypeptide.

Pyrimidine. Một nitrogen base dị vòng có ở trong các nucleotide và nucleic acid.

Pyrophosphate. Phân tử hình thành bởi hai gốc phosphate nối với nhau bằng liên kết anhydride.

Pyrophosphatase. Enzyme thủy phân một pyrophosphate hữu cơ thành hai phân tử orthophosphate.

Quá trình phân tách và tinh sạch đầu ra (downstream processing). Giai đoạn tinh sạch các sản phẩm của một quá trình sinh học (bioprocessing). Các kỹ thuật thường được sử dụng trong bước này là ly tâm, siêu lọc hoặc sắc ký.

Quang tái hoạt hóa (photoreactivation, light repair). Một cách sửa chữa các pyrimidine dimer. Các dimer hồi biến trực tiếp trở lại dạng ban đầu dưới tác dụng của ánh sáng bình thường ở bước sóng 320-370 nm.

Quang hợp (photosynthesis). Quá trình sử dụng năng lượng của ánh sáng để tạo thành carbohydrate bắt đầu từ CO_2 và một tác nhân khử.

Replisome. Một phức hợp đa enzyme có tác dụng khởi động sự tái bản DNA ở chạc ba tái bản.

Retroposon. Loại transposon hoạt động dưới dạng RNA; trước hết DNA được phiên mã thành RNA, RNA này sau đó được phiên mã ngược thành DNA và gắn xen vào một vị trí mới trên hệ gen.

Retrovirus. Là loại virus RNA chứa enzyme reverse transcriptase và sinh sản dưới dạng DNA mạch kép. Chúng có khả năng xâm nhiễm tế bào vật chủ cao. Khi xâm nhiễm nó có khả năng gắn hệ gen của virus với hệ gen của tế bào vật chủ, là cơ sở để thiết kế các vector liệu pháp gen hiệu quả.

Ribonuclease. Enzyme xúc tác đặc hiệu việc phân hủy RNA bằng cách cắt các mối liên kết phosphodiester trên RNA.

Ribonucleic acid (RNA). Thường là phân tử đa phân mạch đơn gồm các đơn vị cấu trúc cơ sở là ribonucleotide. Về mặt hóa học RNA rất giống với DNA. RNA là vật chất di truyền của một số virus và là các phân tử trung gian trong quá trình tổng hợp protein mà thông tin về trình tự amino acid của chúng đã được mã hóa trong DNA.

Ribonucleotide. Đơn vị cấu trúc cơ sở của RNA, gồm ba thành phần: đường ribose, nitrogen base và nhóm phosphate.

Ribosome. Là cơ quan tử khi kết hợp với mRNA tạo ra bộ máy tổng hợp protein. Trong tế bào thường có hàng nghìn ribosome, ribosome của mọi tế bào đều gồm một tiểu đơn vị nhỏ và một tiểu đơn vị lớn. Mỗi tiểu đơn vị có mang nhiều protein và rRNA (trong đó rRNA là thành phần chủ yếu chiếm khoảng 65%) có kích thước khác nhau. Người ta cũng thấy ribosome trong ty thể, ở đó có sự tổng hợp một số protein ty thể.

Ribozyme. Là đoạn RNA có hoạt tính xúc tác phản ứng hóa học giống như enzyme.

RNA bổ sung (complementary RNA). RNA sinh ra bằng cách phiên mã từ khuôn mẫu sợi đơn DNA tương ứng.

RNA kích thước nhỏ của nhân (small nuclear RNA, snRNA). Ngoài mRNA, tRNA Và rRNA, tế bào eukaryote còn chứa những phân tử RNA kích thước nhỏ của nhân (chiếm khoảng <1%) tham gia vào ghép nối các exon.

RNA ligase. Enzyme nối các đoạn RNA với nhau sau khi các intron được cắt rời khỏi tiền chất của mRNA (pre-mRNA) ở các sinh vật eukaryote, tạo ra mRNA hoàn chỉnh sẵn sàng tham gia vào quá trình dịch mã diễn ra trên ribosome.

RNA polymerase. Còn gọi là RNA polymerase phụ thuộc DNA (DNA-dependent RNA polymerase), xúc tác việc tổng hợp RNA trên khuôn mẫu DNA trong quá trình phiên mã.

RNA ribosome (ribosomal RNA, rRNA). Là thành phần cơ bản của ribosome, đóng vai trò xúc tác và cấu trúc trong tổng hợp protein. Tùy theo hệ số lắng rRNA được chia thành nhiều loại: ở eukaryote có rRNA 28S; 18S; 5,8S và 5S; còn các rRNA ở *E. coli* có ba loại: 23S, 16S và 5S. rRNA chiếm nhiều nhất trong bốn loại RNA (khoảng 80% tổng số RNA tế bào), tiếp đến là tRNA khoảng 16%, mRNA chỉ khoảng 2%. Ngoài ra, tế bào sinh vật eukaryote còn chứa những phân tử RNA kích thước nhỏ của nhân (small nuclear, snRNA) chiếm khoảng <1% tham gia vào ghép nối các exon.

RNase. Enzyme thủy phân RNA.

RNA thông tin (messenger RNA, mRNA). Một loại RNA được phiên mã từ một trình tự DNA. mRNA truyền thông tin di truyền từ nhiễm sắc thể tới ribosome để tổng hợp protein. Trong quá trình đó một sợi của chuỗi xoắn kép DNA được dùng làm khuôn mẫu, dọc theo nó các nucleotide của mRNA bổ sung được xếp thành hàng, nối với nhau tạo nên một polynucleotide giống hệt sợi DNA không làm khuôn mẫu ngoại trừ thymine được thay bằng uracil. Quá trình này gọi là phiên mã và phân tử mRNA mang mã di truyền được dùng để điều khiển sự hình thành protein trên ribosome.

RNA vận chuyển (transfer RNA, tRNA). Loại RNA mang các amino acid đến ribosome và sắp xếp chúng dọc theo phân tử mRNA đã nằm sẵn ở đó. Tại đây, các amino acid nối với nhau bằng liên kết peptide để tạo thành phân tử protein. Mỗi amino acid có

một phân tử RNA vận chuyển riêng với bộ ba đặc trưng và như vậy các amino acid được sắp xếp theo trật tự của các nitrogen base trên mRNA trong quá trình dịch mã.

10 sequence. Là đoạn mã hóa có tác động liên ứng TATAATG tập trung khoảng 10 bp trước điểm khởi đầu của một gen của vi khuẩn.

35 sequence. Là đoạn mã hóa tập trung khoảng 35 bp trước điểm khởi đầu của một gen của vi khuẩn.

Sản phẩm phiên mã sơ cấp (primary transcript). Sản phẩm phiên mã ban đầu của các gen eukaryote còn được gọi là mRNA tiền thân (pre-mRNA), thường có kích thước rất lớn. Nó phải được xử lý (splicing) để tạo ra phân tử mRNA hoàn chỉnh (mature mRNA) dùng cho dịch mã.

Sàng lọc (screening). Kỹ thuật nhận dạng một dòng DNA trong một thư viện hệ gen (genomic library) hoặc thư viện cDNA (cDNA library) bằng một phương pháp lai mẫu dò có đánh dấu [α-^{32}P]dCTP với các vết tan (trường hợp dùng bacteriophage λ làm vector tạo dòng và cho xâm nhiễm vào vi khuẩn *E. coli*) hoặc khuẩn lạc (dùng plasmid làm vector tạo dòng) của các thư viện đó trên màng nylon hoặc nitrocellulose. Tín hiệu lai được phát hiện bằng phóng xạ tự ghi trên phim X-quang.

Sinh học phân tử (molecular biology). Khoa học nghiên cứu các hiện tượng sống ở mức độ phân tử. Lĩnh vực khoa học trẻ tuổi này là điểm gặp nhau của các khoa học kinh điển như di truyền học, hóa sinh học, tế bào học, vật lý học, hóa học hữu cơ và hóa lý. Theo cách hiểu phổ biến hiện nay, sinh học phân tử là khoa học nghiên cứu các gen và hoạt động của chúng ở mức độ phân tử, bao gồm phiên mã, dịch mã, sao chép, điều hòa biểu hiện gen, tái tổ hợp và chuyển gen...

Sinh tổng hợp protein (protein synthesis). Phản ứng hóa học diễn ra trên ribosome tạo nên các phân tử protein từ các amino acid trên cơ sở thông tin di truyền nhận được từ trong nhân tế bào thông qua mRNA.

Somatotropin. Đây là loại hormone đặc biệt, thường được tổng hợp trong não động vật và người với một hàm lượng vô cùng thấp. Somatostatin có vai trò điều hòa hormone sinh trưởng và insulin đi vào máu, kiểm tra sự tổng hợp hai loại hormone này. Hiện nay, hormone này đã được sản xuất bằng công nghệ DNA tái tổ hợp để chữa bệnh lùn ở trẻ em do thiếu somatotropin.

Southern blot. Kỹ thuật chuyển và cố định DNA đã biến tính từ agarose gel (sau khi được phân đoạn bằng điện di) lên màng lai bằng nylon hay nitrocellulose để lai với mẫu dò (probe) được đánh dấu đồng vị phóng xạ [α-^{32}P]dCTP hoặc digoxigenin-dUTP. Tín hiệu lai sau đó được phát hiện trên phim X-quang (trường hợp [α-^{32}P]dCTP) hoặc trên màng lai (trường hợp digoxigenin-dUTP).

Số bản sao (copy number). (1) Số các phân tử plasmid có trong một tế bào vi khuẩn. (2) Số lượng các bản sao của một gen trong hệ gen của một sinh vật.

Sơ đồ phóng xạ tự ghi (autoradiogram). Hình ảnh sinh ra trên phim X-quang do sự phát xạ của các hạt phóng xạ.

Sợi chủ (leading strand). Chuỗi DNA trong quá trình tái bản được tổng hợp cùng chiều với hướng di chuyển của chạc ba tái bản.

Sợi có nghĩa (sense strand). Sợi đơn trong chuỗi xoắn kép DNA được dùng làm khuôn mẫu để tổng hợp mRNA trong quá trình phiên mã. Sợi đơn DNA thứ hai không được dùng làm khuôn mẫu để phiên mã gọi là sợi đối nghĩa (antisense strand).

Sợi thứ (lagging strand). Chuỗi DNA trong quá trình tái bản được tổng hợp ngược chiều với hướng di chuyển của chạc ba tái bản.

Suppressor. Có hai loại: (1) Extragenic-thường là một gen mã hóa một tRNA đột biến có trên codon đột biến, theo nghĩa của codon đầu tiên. (2) Intragenic-là đột biến có tính chất bù đắp để phục hồi khung đọc mã đầu tiên sau khi đã có sự chuyển dịch khung.

Sửa chữa bằng cắt bỏ (excision repair, dark repair). Quá trình sửa chữa các pyrimidine dimer trên DNA (do tia tử ngoại sinh ra) bằng cách cắt bỏ các dimer và tổng hợp đoạn DNA mới thay thế nào đó dưới tác dụng của các enzyme đặc hiệu mà không cần đến ánh sáng.

Tác nhân đột biến (mutation factor). Bất kỳ tác nhân vật lý hoặc hóa học nào làm tăng đáng kể tần số đột biến so với tần số đột biến tự nhiên thì đều được xem là tác nhân đột biến. Các tác nhân đột biến thông dụng là các tia phóng xạ như tia X, tia gamma, tia tử ngoại; các hóa chất như các đồng đẳng của các nitrogen base, nitrogen acid, hydroxylamine, acridin...

Tái bản (replication). Sự sao chép vật chất di truyền trong chu trình phân bào hoặc sự tổng hợp DNA của phage khi phage sinh sản trong tế bào vi khuẩn.

Tái bản DNA kiểu bán bảo toàn (semiconservative DNA replication). Quá trình sao chép DNA mà hai sợi cha mẹ dãn xoắn và được dùng làm khuôn để tổng hợp hai sợi mới. Khi sao chép xong thì mỗi chuỗi xoắn kép con gồm một sợi cũ và một sợi mới.

Tái bản DNA kiểu gián đoạn (discontinuos DNA replication). Sự tổng hợp DNA thành những đoạn ngắn, sau đó nối lại với nhau tạo thành chuỗi polynucleotide dài.

Tái bản DNA kiểu nửa gián đoạn (semidiscontinuos DNA replication). Sự tổng hợp DNA liên tục trên một sợi khuôn mẫu trong hai sợi của DNA mạch kép; còn trên sợi khuôn mẫu kia, DNA được tổng hợp thành những đoạn ngắn gọi là đoạn Okazaki, sau đó tất cả chúng được nối lại với nhau nhờ ligase tạo thành chuỗi polynucleotide dài.

Tái bản hai hướng (bidirectional replication). Sự tái bản theo hai hướng ngược chiều nhau.

Tái tổ hợp (recombination). Quá trình mà trong đó nhiễm sắc thể hay phân tử DNA đứt ra rồi các phần đứt được nối lại theo một tổ hợp mới. Quá trình này có thể xảy ra trong tế bào sống (qua sự trao đổi chéo trong phân bào giảm nhiễm) hay trong ống nghiệm nhờ các enzyme cắt và nối DNA.

Tạo dòng (cloning). Còn gọi là nhân dòng, tách dòng hay dòng hóa, là sự sản sinh nhiều bản sao của một phân tử DNA, thường là phân tử DNA tái tổ hợp trong plasmid vector, bằng cách sao chép phân tử đó trong một vật chủ thích hợp chẳng hạn vi khuẩn *E. coli*.

Tần số lặp lại (repetition frequency). Là số lần sao chép của một đoạn mã hóa nào đó có trong hệ gen đơn bội, tương đương với 1 trong trường hợp nonrepetitive DNA và lớn hơn 2 trong trường hợp repetitive DNA.

Terminal transferase. Enzyme bổ sung các gốc nucleotide vào đầu 3' của oligonucleotide hoặc polynucleotide.

Tế bào bạch huyết (B lymphocyte). Là những tế bào có chức năng tổng hợp các kháng thể.

Tế bào Hfr (high frequency recombination cell). Tế bào giới tính đực ở *E. coli*, có mang nhân tố F gắn liền với nhiễm sắc thể vi khuẩn. Khi nhân tố F thúc đẩy sự tiếp hợp của tế bào Hfr với tế bào cái (F^-) thì các gen của vi khuẩn được truyền sang tế bào cái với tần số cao.

Tế bào mầm phôi (embryonic stem cell). Tế bào phôi chưa biệt hóa, đa thể ở chuột, có thể được nuôi cấy trong một thời gian dài mà vẫn giữ được tính đa thể, nghĩa là khả năng biệt hóa theo nhiều hướng để phát triển thành các tế bào khác nhau như: tế bào cơ tim, tế bào thần kinh, tế bào gan...

Thể biến nạp (transformant). Tế bào hoặc sinh vật nhận được gen của một sinh vật khác trong quá trình biến nạp và biểu hiện chức năng của gen đó ra kiểu hình.

Thể đa hình (polymorphism). Mô tả sự có mặt đồng thời của quần thể trong hệ gen biểu hiện biến dị có tính chất allele, có thể quan sát trên các allele tạo ra các kiểu hình khác nhau, hoặc sự thay đổi của DNA ảnh hưởng đến kiểu cắt hạn chế (restriction patterns).

Thể đột biến (mutant). Sinh vật (hoặc gen) mang đột biến di truyền.

Thể khảm (mosaic). Phôi hoặc cơ thể có các tế bào mang các hệ gen không giống nhau.

Thể nhân (nucleosome). Cấu trúc cơ sở của các nhiễm sắc thể eukaryote gồm DNA bao quanh một lõi histone. Thể nhân có dạng hạt, quan sát thấy dưới kính hiển vi điện tử.

Thể tái tổ hợp (recombinant). Các cá thể hoặc tế bào mang các tổ hợp gen khác với cha mẹ của chúng do các quá trình tái tổ hợp di truyền sinh ra.

Thông tin di truyền (genetic information). Thông tin được lưu trữ trong các phân tử DNA của sinh vật ở dạng trình tự sắp xếp của bốn nucleotide ký hiệu là A, T, C và G đóng vai trò như những "chữ cái" của "ngôn ngữ" di truyền. Trong ngôn ngữ này, mỗi từ chỉ có ba chữ cái gọi là một bộ ba. Nghĩa của mỗi từ là một amino acid có mặt trên phân tử protein tương ứng. Mỗi "câu" của ngôn ngữ di truyền là một gen chứa đựng thông tin

di truyền để đảm nhiệm một chức năng trọn vẹn. Mỗi chức năng là một đặc tính sinh lý, hình thái hay cấu trúc của sự sống. Do cơ chế sao chép theo kiểu nửa bảo toàn của DNA mà thông tin di truyền được truyền chính xác từ thế hệ nọ sang thế hệ kia hầu như không thay đổi.

Thụ thể (receptor). Protein màng, một phần nằm trong màng tế bào, có khả năng gắn một phối tử (ligand) ở mặt ngoài của màng kết quả là gây ra một biến đổi hoạt động trong tế bào chất.

Thuốc thử nghiệm (orphan drug). Là thuốc được pháp luật Mỹ (và một số nước khác) cho phép dùng thử trong một vài năm (ở Mỹ là 7 năm) để chữa những căn bệnh hiếm gặp. Hormone phát triển người và erythropoetin được sản xuất nhờ công nghệ DNA tái tổ hợp thuộc dạng thuốc thử nghiệm nói trên.

Thư viện cDNA (cDNA library). Tập hợp các dòng DNA được tạo ra từ mRNA của một tế bào hoặc một mô cụ thể trong bacteriophage vector, đại diện cho thông tin di truyền mà các tế bào đó biểu hiện.

Thư viện hệ gen (genomic library). Tập hợp tất cả các đoạn DNA được tạo ra từ phản ứng cắt hạn chế genome trong bacteriophage vector, đại diện được cho toàn bộ cho thông tin di truyền của một hệ gen.

Thymine (T). Base loại pyrimidine có trong DNA nhưng không có trong RNA. Trong mạch kép DNA thymine bắt cặp với adenine (A).

Tính trạng (trait). Một đặc điểm hoặc tính chất bên ngoài của các cá thể sinh vật.

Tính trạng đa gen (polygenic trait). Tính trạng được xác định bởi mối tương tác giữa nhiều gen, chẳng hạn màu mắt ở người.

Topoisomerases. Những enzyme xúc tác sự thay đổi những dạng siêu xoắn âm hoặc dương của chuỗi DNA xoắn kép vòng.

Trao đổi chất (metabolism). Còn gọi là quá trình chuyên hóa, là toàn bộ những thay đổi trong tế bào sống của những phân tử hữu cơ nhờ xúc tác bởi những enzyme. Sự chuyển hóa bao gồm cả quá trình tổng hợp lẫn thoái hóa.

Trình tự dẫn đầu (leader sequence). Một trong ba phần chủ yếu của một phân tử mRNA. Trình tự này nằm ở đầu 5’ của mRNA và mang thông tin để ribosome và các protein đặc hiệu nhận biết bắt đầu quá trình tổng hợp polypeptide, trình tự dẫn đầu không được dịch mã thành trình tự các amino acid.

Trình tự điều hòa (regulatory sequence). Một trình tự của DNA tham gia vào quá trình điều hòa của gen. Ví dụ: trình tự promoter hoặc operator.

Trình tự chèn (insertion sequence, IS). Có ở vi khuẩn, kích thước nhỏ và tương đương với transposon, chỉ mang gen cần cho sự gắn chèn của chính nó.

Trình tự chỉ huy (operator). Đoạn DNA ngắn nằm ở đầu operon, kề sát promoter nơi mà protein ức chế có thể bám vào, có tác dụng “mở” hoặc “đóng” operon để cho các gen cấu trúc trong operon hoạt động hoặc ngừng hoạt động.

Trình tự khởi động (promoter). Trình tự nucleotide đặc hiệu nằm trong thành phần operon, có chức năng điều hòa hoạt động của operon, nơi RNA polymerase bám vào để bắt đầu quá trình phiên mã. Trình tự đặc trưng của promoter có khoảng 20-200 nitrogen base.

Trình tự lặp lại (repetitive sequence). Các trình tự base có nhiều bản sao cùng một lúc trên phân tử DNA. Có bốn loại trình tự lặp lại: trình tự lặp lại đơn (có một bản sao), trình tự lặp lại vừa (có từ 1 đến 10 bản sao), trình tự lặp lại trung bình (có từ 10 đến vài trăm bản sao) và trình tự lặp lại cao (có từ vài trăm đến vài triệu bản sao).

Trình tự lặp lại ngược chiều (inverted repeat sequence). Một đoạn ngắn của DNA được lặp lại, thường ở các đầu của một đoạn dài hơn, có hướng ngược chiều.

Trình tự lặp lại xuôi chiều (direct repeat sequence). Đoạn DNA gồm một dãy các trình tự nucleotide được lặp lại theo cùng một chiều.

Trình tự phổ biến (consensus sequence). Đoạn trình tự thấy có trong hầu hết các mẫu của một yếu tố di truyền cụ thể và có độ bền vững cao, ví dụ hộp CAAT.

Trình tự Shine-Dalgarno (Shine Dalgarno sequence, SD). Còn gọi là vùng liên kết ribosome (RBS), là một phần của trình tự nucleotide ở đầu 5' của một mRNA prokaryote có thể kết hợp bổ sung cặp base với đầu 3' của 16S rRNA, dùng làm tín hiệu cho sự khởi đầu dịch mã.

Trình tự tăng cường (enhancer). Trình tự nucleotide dạng *cis* làm tăng cường độ phiên mã của promoter trong gen eukaryote. Nó có thể nằm cách promoter hàng ngàn cặp base và hoạt động theo cả hai hướng ở bất kỳ vị trí nào so với promoter.

Tương đồng (homologous). (1) Nói về các nhiễm sắc thể cặp đôi trong các sinh vật lưỡng bội. (2) Dùng để mô tả các trình tự DNA giống hệt nhau một cách tuyệt đối; tuy nhiên, phần trăm tương đồng giữa các đoạn liên quan đôi khi chỉ tương đối.

Ty thể (mitochondria). Bào quan có màng bao bọc, nằm trong tế bào chất của các sinh vật eukaryote. Ty thể chứa những enzyme cần thiết cho chu trình citric acid, vận chuyển electron và quá trình phosphoryl oxy hóa.

Ủ để gắn mồi (annealing). Dùng để chỉ sự bắt cặp của khuôn mẫu DNA sợi đơn với primer (mồi) để tổng hợp nên một sợi DNA bổ sung bằng cách sử dụng các dNTP có trong môi trường để kéo dài primer nhờ sự xúc tác của enzyme Taq DNA polymerase (trong khuếch đại PCR) hoặc DNA polymerase I (trong tổng hợp cDNA).

Ức chế amber (amber suppresser). Là các gen đột biến (mã hóa cho tRNA) mà những anticodons của nó đã được kích hoạt để có thể nhận UAG codon cũng như các codon trước đó.

Vật chủ (host). Tế bào dùng để nhân các phân tử DNA lên nhiều lần.

Vector. Là các phân tử DNA được sử dụng trong tạo dòng và biểu hiện gen, và nhân bản chúng trong tế bào vật chủ (*E. coli* hoặc nấm men). Có ba nhóm vector chính gồm: (1) Nhóm plasmid, (2) Nhóm phage/phagemid, và (3) Nhóm nhiễm sắc thể nhân tạo

(artificial chromosome: BAC, YAC…). Ý tưởng về vector chuyển gen bắt nguồn từ các plasmid của vi khuẩn. Vector chuyển gen là các phân tử DNA có khả năng tự tái sinh, tồn tại độc lập trong tế bào và mang được các gen cần chuyển.

Vector biểu hiện (expression vector). Là vector có thể mang các gen ngoại lai mong muốn cho phép thực hiện sự phiên mã của các bản sao được tạo dòng và sự dịch mã các mRNA của chúng trong tế bào vật chủ, ví dụ: *E. coli*. Để biểu hiện các gen ngoại lai trong *E. coli* phải bắt đầu bằng việc gắn nó vào trong vector biểu hiện (thường là plasmid). Vector này phải có đủ các cấu trúc cần thiết sau: (1) Các trình tự mã hóa gen chỉ thị chọn lọc (selectable marker) để đảm bảo duy trì vector trong tế bào. (2) Một promoter kiểm soát phiên mã cho phép sản xuất một lượng lớn mRNA từ các gen được tạo dòng. (3) Các trình tự kiểm soát dịch mã như vùng liên kết ribosome được bố trí thích hợp và codon khởi đầu ATG. (4) Một polylinker để đưa gen ngoại lai vào trong một hướng chính xác với promoter.

Vector tạo dòng (cloning vector). Phân tử DNA mạch kép có khả năng tự sao chép trong tế bào vật chủ. Có thể gắn vào phân tử này một đoạn hoặc một vài đoạn DNA khác nguồn tạo nên phân tử DNA tái tổ hợp dùng để nhân dòng.

Vector thay thế (replacement vector hay **substitution vector).** Là một bacteriophage vector mà ở đó các điểm tạo dòng sắp xếp thành các cặp, do vậy phần của hệ gen nằm giữa các điểm này có thể được thay thế bằng đoạn DNA chèn.

Vết tan (plaque). Vòng tròn trong suốt xuất hiện trên thảm đục của các vi khuẩn mọc trên môi trường thạch đặc, do sự tan vỡ lặp lại nhiều chu kỳ của các tế bào vi khuẩn bị bacteriophage xâm nhiễm và sinh tan.

Vị trí *cos* (*cos* site). Vùng được tạo ra khi các đầu dính của DNA phage λ nối lại với nhau.

Virus. Phức hợp chứa nucleic acid (DNA hoặc RNA) nằm trong một vỏ bọc protein, có khả năng gây nhiễm và tái bản bên trong tế bào vật chủ đặc hiệu, tạo ra nhiều virus, lan truyền từ tế bào này sang tế bào khác. Virus là dạng sống không có cấu trúc tế bào, có khả năng xâm nhập vào các tế bào sống xác định và chỉ sinh sản ở bên trong các tế bào đó. Giống như tất cả cá sinh vật khác, virus có bộ máy di truyền của riêng mình, mã hóa việc tổng hợp các hạt virus từ các chất có trong tế bào vật chủ. Như vậy, virus là những vật ký sinh nội bào. Virus phân bố ở khắp nơi trong tự nhiên, xâm nhập vào tất cả các nhóm sinh vật. Người ta đã biết khoảng 500 loại virus xâm nhập động vật máu nóng, 300 loại xâm nhập thực vật bậc cao. Một số khối u ung thư ở động vật và ở người có thể do virus. Virus tồn tại ở hai dạng: dạng nghỉ hay ngoại bào và dạng sinh sản hay nội bào. Kích thước của các hạt virus từ 15-350 nm, chiều dài của một số loại virus có thể đạt tới 2000 nm. Phần lớn virus chỉ nhìn thấy được qua kính hiển vi điện tử. Chất mang thông tin di truyền của virus là nucleic acid: DNA hoặc RNA. Vì vậy, có thể phân virus thành hai loại: loại mang DNA và loại mang RNA.

Vi tiêm (microinjection). Kỹ thuật đưa DNA vào nhân hoặc vào tế bào chất của tế bào bằng kim mao dẫn và bơm áp lực. Đây là kỹ thuật sử dụng phổ biến trong công nghệ tế bào động vật (animal cell biotechnology), nó đòi hỏi thiết bị vi thao tác (micromanipulator) cực nhạy, kỹ năng thực hành và sự kiên nhẫn cao của kỹ thuật viên.

VNTR (variable number of tandem repeats). Còn gọi là tiểu vệ tinh, chỉ các vùng của hệ gen đặc trưng bởi sự lặp lại của cùng một trình tự DNA. Số lượng lần lặp lại thay đổi theo cá thể nên mang tính đa allele rất cao.

Vòng kẹp tóc (hairpin loop). Vùng chuỗi đơn bổ sung tạo nếp gấp chứa các cặp base tạo thành xoắn kép, được dùng làm mồi (primer) cho quá trình tổng hợp sợi cDNA thứ hai.

Vốn gen (gene pool). Toàn bộ thông tin di truyền có trong tất cả các gen của một quần thể tại một thời điểm xác định.

Vùng cùng hướng (downtream). Đề cập đến vị trí của một đoạn trình tự nào đó nằm ở phía đầu 3' của gen hoặc một đoạn gen quan tâm.

Vùng đa nối (polylinker hay **polycloning sites).** Một trình tự DNA mạch kép được tổng hợp nhân tạo có mang một loạt các vị trí nhận biết của các enzyme hạn chế. Trình tự này được gắn vào các vector dùng trong kỹ thuật tạo dòng gen (như vùng tạo dòng).

Vùng liên kết ribosome (ribosome binding sites, RBS). Vùng trên phân tử mRNA giúp cho nó bám vào ribosome trong quá trình dịch mã (xem trình tự Shine-Dalgarno).

Vùng nhân (nucleoid). Một vùng trong tế bào vi khuẩn tập trung vật chất di truyền.

Vùng ngược hướng (upstream region). Vị trí của một trình tự nucleotide nào đó nằm ở phía đầu 5' của phân tử DNA so với gen quan tâm.

Vùng siêu biến (hypervariable region, HVR). Vùng trong hệ gen bao gồm một số lượng biến động các trình tự lặp lại và được dùng để xác định đặc trưng cá thể (xem in dấu di truyền).

Vùng tạo dòng (multiple cloning sites, MCS). Đoạn DNA ngắn trong một vector thế hệ thứ ba (ví dụ: các nhóm pUC, pGEM, pBluescript…) có chứa các vị trí nhận biết cho một số enzyme cắt hạn chế thông dụng, được thiết kế để chèn đoạn DNA ngoại lai vào đây (xem vùng đa nối).

Western blot. Kỹ thuật phân tích protein dựa trên nguyên lý liên kết kháng nguyên-kháng thể để phát hiện protein đặc hiệu có bản chất kháng nguyên (thường được thẩm tích lên màng nitrocellulose sau khi chạy điện di SDS-PAGE, và cố định ở đó). Sau khi protein trên màng lai gắn với kháng thể thứ nhất đặc hiệu và tiếp đến là kháng thể thứ hai có đánh dấu enzyme (alkaline phosphatase hoặc horse-radish peroxidase…) thì phức hợp này sẽ được liên kết với cơ chất để tạo màu. Sự hiện diện của protein ngoại lai (sản

phẩm dịch mã của gen ngoại lai được chuyển vào tế bào vật chủ) sẽ được phát hiện nhờ sự xuất hiện màu của phản ứng lai.

Xoắn alpha (helix r). Cấu trúc xoắn của chuỗi polypeptide, thường là quay phải, với số lượng tối đa liên kết trong chuỗi, một trong những cấu trúc bậc hai phổ biến của protein được tìm thấy ở keratin.

Xoắn kép (double helix). Cấu trúc ba chiều tự nhiên của hai chuỗi DNA bổ sung, đối song và xoắn với nhau.

YAC (Yeast artificial chromosome). Nhiễm sắc thể nhân tạo của nấm men, được dùng làm vector để tạo dòng những đoạn DNA có kích thước rất lớn trong nấm men.

Yếu tố ngoài nhiễm sắc thể (extrachromosomal element). Phân tử DNA không phải là thành phần của nhiễm sắc thể tế bào vật chủ.

Yếu tố tác động *cis* (*cis*-acting element). Đoạn trình tự DNA chỉ biểu hiện hiệu quả trên chính phân tử DNA mà nó tác động. Ví dụ: hộp CAAT là một phần tử tác động *cis* đối với quá trình phiên mã ở các sinh vật eukaryote.

Yếu tố tác động *trans* (*trans*-acting element). Yếu tố di truyền có thể biểu hiện hiệu quả mà không cần nằm trên cùng một phân tử với đoạn trình tự đích. Thường yếu tố đó mã hóa cho một sản phẩm protein (có thể là một enzyme hay một protein điều hòa) và sản phẩm này có thể khuếch tán đến điểm tác động.

Tài liệu tham khảo/đọc thêm

1. Ban Từ điển-NXB Khoa học và Kỹ thuật. 2002. Từ điển Bách khoa Sinh học. *NXB Khoa học và Kỹ thuật*, Hà Nội.

2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. 2002. Molecular Biology of the Cell. 4th ed. *Garland Science Publishing, Inc.* New York, USA.

3. Karp G. 2002. Cell and Molecular Biology. *John Wiley & Sons, Inc.* New York, USA.

4. Lawrence E. 1995. Henderson's Dictionary of Biological Terms. 7th ed. *Longman Group Ltd*. Singapore.

5. Singleton P and Sainsbury D. 2001. Dictionary of Microbiology and Molecular Biology. 3rd ed. *John Wiley & Sons, Ltd*. UK.

6. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Losick R. 2004. Molecular Biology of the Gene. *Pearson Education, Inc./Benjamin Cummings Publishing*, San Francisco, USA.